THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY

VOLUME XVIII, 1966

PUBLISHED FOR
The Experimental Psychology Society
BY
W. Heffer & Sons Ltd. Cambridge

The Quarterly Journal of Experimental Psychology

Volume XVIII

CONTENTS

SPECIAL ARTICLES

SHORTER ARTICLES AND NOTES

APPARATUS

THE QUARTERLY JOURNAL OF
EXPERIMENTAL PSYCHOLOGY

Vol. XVIII 1966 Part 1

THE PERFORMANCE OF A SIMPLE MOTOR TASK WITH KINAESTHETIC SENSE LOSS

BY

JUDITH I. LASZLO

From the University of Western Australia

Two related studies were carried out, to test the suitability of the nerve compression block as a technique in the investigation of kinaesthesis in motor skills.

In both studies a key tapping task was used. Each experimental group was composed of six volunteer subjects. It was found that kinaesthetic sensation was eliminated after pressure had been applied for 20–25 min., but muscle power was not seriously affected at this stage of the block. The results also showed a pronounced performance decrement in the absence of kinaesthetic feedback, and that this decrement was not due to emotional or other disturbances caused by the experimental procedure. The loss of tactile sensation was also observed.

Introduction

Skilled performance of a motor task depends to a great extent on information received by the organism, through its various sense organs, as to the progress of the response already initiated. The sense modalities which can play a role in the monitoring process include exteroceptive and proprioceptive sensations. In the former group visual, auditory and tactile feed-backs can be mentioned, while kinaesthetic sensation is the relevant representative of the latter group.

Most authors working in the field of motor skill monitoring mention kinaesthesis as an important variable. In a comprehensive review of the relevant literature (Adams and Creamer, 1962), Adams points out that the hypothesized role of kinaesthesis in motor performance is a regulatory one aiding the organism in discriminating the correct response from the incorrect one. In this way kinaesthesis supplements the feedback provided by the exteroceptive senses.

There have been attempts at reduction of kinaesthetic feedback. Chase *et al.* (1961) placed vibrators on the arm, with the subject performing a keytapping task. Chase says that "We felt that vibration on the arm would distort the decoding of the proprioceptive information by the central nervous system . . ." (p. 154). He could not, however, prove this contention, hence it is questionable how far kinaesthetic feedback was, in actual fact, reduced.

Provins (1958), in a study mainly concerned with the different feedback mechanisms subserving passive and active movement includes the investigation of rate decrement in isometric finger oscillation under xylocaine nerve block. The xylocaine was injected around the metacarpophalangeal joint. He found a small (6 per cent.), but statistically significant, decrease in oscillation rate. In this situation, however, only the skin and joint receptors were anaesthetized while muscle and tendon receptors responsible for the movements involved in finger oscillation were intact.

Hence, Provins's work shows a decrement in performance with selectively reduced feedback, rather than with a general cut-out of feedback.

A search for a possible technique for the more nearly complete reduction of kinaesthesis led the author to investigate the possibility of using "nerve compression locks" (Sinclair and Henshaw, 1951; Sinclair, personal communication, 1963).

This technique can be described briefly as follows: a sphygmomanometer is applied to the upper arm at a pressure exceeding the subject's blood pressure. This way the blood supply to the limb distal to the cuff is cut off. The reduced O_2 supply affects the nerve fibres, inducing the typical block syndrome. In these block syndromes various sense modalities and muscle power tend to be reduced progressively and are eventually abolished. This elimination of the various sense modalities is said to follow a set order. If kinaesthetic sensation cuts out at a reasonably early stage, and if it becomes eliminated before appreciable motor impairment sets in, the compression block method seems a distinct experimental possibility in the control of proprioception in motor skills.

Nerve compression blocks, to the best knowledge of the author, have not been used in the psychological investigations of motor skills as yet. In the two studies here reported, the author attempted to test the suitability of the nerve compression block technique in the investigation of kinaesthesis in motor skills. As an untried area, it presents many questions and problems. The two studies were conducted to answer and clarify these. Both could be classified as exploratory work, in the "see what happens" category.

The questions posed were:

(1) When does kinaesthetic sensation cut out under compression block?
(2) How is kinaesthetic sense related to touch in time sequence of elimination?
(3) Is there a span of time between loss of kinaesthesis and motor impairment?
(4) Is is possible to perform a motor skill with the cuff on?
(5) If decrement in performance is found with the cuff on, is this due to lack of kinaesthetic feedback, or some other factor, e.g. motor impairment, emotional disturbance?

Procedure

(a) *The compression block.** Subject's blood pressure is taken, and the pressure from the cuff is lowered to zero. Leaving the cuff in position, midway between shoulder and elbow joint, the subject's arm is raised into the vertical position, and the pressure in the cuff is suddenly raised to approximately 40 mm. Hg, higher than the systolic pressure. The arm is then lowered into the horizontal position. This pressure head is necessary to ensure complete stoppage of the blood supply, even with a possible, emotionally caused rise in blood pressure. The pressure is then held on that point throughout the experiment. The subject's arm is resting on a soft cushion, and is covered with woollen material, to avoid a drop in skin temperature. When testing is completed, pressure is gradually lowered, the cuff removed. The subject rests his arm for 2–3 min. until the sensation of "pins-and-needles" disappears. Throughout the testing session, except for the actual tapping, the arm and hand are lying quietly, fingers only moving at the request of the experimenter.

(b) *Testing tactile, passive and active kinaesthetic senses.* At regular intervals, with the subject closing his eyes, and head turned away, the following tests are performed on thumb or any one finger-tip: (1) touched lightly with a cotton-wool-covered small wooden stick (Johnson's cotton buds); (2) moved up or down; (3) subject is asked to move it up or down. (2) and (3) involved identical movements. It was found that, from the commencement of the block, the test can be performed in the tenth minute, then every second minute to sixteenth minute, and from then on at every minute. The three tests were performed in random order, and varying the fingers from test to test.

Subject verbally reports when and where he feels touch or movement, or when he is moving the finger.

* Compression block technique should not be used without medical supervision. It is a safe technique if applied by an experimenter trained in physiology or medicine. However, overlong application of pressure, or daily exposure could cause possible nerve injury. The choice of subjects has to be carefully considered also, excluding subjects with circulatory deficiencies.

Experiment 1. Six subjects, all volunteers (five males, one female) were tested individually. (One of the male subjects took part in both Expts. 1 and 2.) At the beginning of each experimental session the experimenter outlined the procedure, showed the subject the apparatus, and answered questions, if any. Subject was then asked to tap on a morse key at his fastest rate, as evenly as possible. The tapping, and a 1 sec. time signal, were recorded by a two-pen incident recorder. Approximately 30–40 taps were given. The compression block was then applied, and sense loss tested. When both passive and active kinaesthetic sense became eliminated, the subject was asked to tap in the same way as before with the cuff on. Lastly, the subject was asked to clasp the experimenter's hand and squeeze it as hard as he was able, still with the cuff on. After this the pressure was lowered and the session terminated.

Experiment 2. Subjects (six volunteers, five males, one female) tapped on the key first with preferred, then with non-preferred, hand for 20 sec. each. Cuffs were then applied to both arms. On the non-preferred hand sensory tests were performed once only as soon as subject reported the cessation of "pins-and-needles" (approximately 6–7 min. after pressure application). Subject was then asked to tap as before with the non-preferred hand for 20 sec., and then the pressure was released. The preferred hand was treated as in Expt. 1, until sensations cut out. 20 sec. tapping and hand-squeezing followed. Pressure was released and session ended.

Apparatus

Two sphygmomanometers, stethoscope, two stopwatches, morse-key connected to recording apparatus. On the record strip of the tapping a 1 sec. time signal was also recorded. Cushion, small woollen rug, cotton buds.

Results

Sensory loss data from both Experiments 1 and 2 are treated together.

TABLE I

Disappearance of the Sensations for Each Individual Subject from the Commencement of the Compression Block

Subject	*Tactile (min.)*	*Kinaesthetic (min.)* Passive	*Active*
1	17, 22	22	21
2	20	23	23
3	18	21	22
4	22, 24	24?	22
5	18	22	22
6	19	21–23	24
6	17, 22	21	21
6	16	20?	18
7	20, 22	23	23
8	22, 24	22, 24	24
9	21	23	25
10	20	22?	24
11			
Mean	19·2	22	22·4
σ	1·75	1·13	1·88

S6 took part in both experiments.
? denotes errors or uncertainty in verbal report.

For tactile and passive kinaesthetic sensation the figures listed above denote the times when sensation first disappeared. In three subjects tactile sensation reappeared for approximately 1 min. and then finally disappeared. In a fourth

subject a 5 min. reappearance was observed. For passive kinaesthetic sense, two 1 min. and one 3 min. reappearance of the sensation was recorded.

Active kinaesthesis, though following the final cut-out time of passive kinaesthesis closely in all subjects, is sudden and well-defined.

A difference in cut-off times on the two occasions of testing for S6 can be observed.

Muscle impairment was estimated by the hand squeeze. Ten of the subjects were able to squeeze the experimenter's hand with appreciable strength. In one subject the pressure was very weak, and one subject refused to try, saying that he couldn't "feel" his hand. Both these subjects were able to perform the tapping, though at a much reduced rate as compared to pre-cuff records.

Comparison of tapping performance in normal and compression block condition may now be examined.

Experiment 1

Table II shows a great decrease in the number of taps for each 5 sec. period after the loss of kinaesthesis as compared to the pre-cuff performance. Not only was tapping slow, but it showed great irregularity in rhythm and in the length of time the key was held depressed.

TABLE II

Experiment 1

Number of Taps for Each Subject in 5 sec. Intervals

Condition	*Subject*	*S1*	*S2*	*S3*	*S4*	*S5*	*S6*
	sec.						
Precuff	0–5	24	19	30	20	28	12
	5–10	30	24	29	23	30	13
Mean ..	.. ..	27	21·5	29·5	21·5	29	12·5
Cuff	0–5	3	7	recording apparatus broke down, S was slowly tapping	15	3	2
	5–10	7	3		17	5	7
	10–15	11	–*		17	–*	4
	15–20	–*	–		15	–	9
Mean ..	.. ..	7	5		16	4	5·5

* Stopped work.

Another notable point is, that in all but one subject (S2) the second 5 sec. unit shows an improved score. Further increase in number of taps is observable in S1 and S6.

S4, however, shows no irregularity but only decreased speed. During the testing of S3, the Morse-key recording apparatus broke down, but S was observed to tap slowly. Incidentally, S3 was the one who showed a weak hand-squeeze response. For the other five subjects, hand squeeze was fairly strong.

Experiment 2

The results in Table III show that performance does not deteriorate with pressure applied to the arm prior to kinaesthetic sense loss.

TABLE III

EXPERIMENT 2

PERFORMANCE SCORED IN NUMBER OF TAPS IN 5 SEC. INTERVALS

		Subjects	*S6*	*S7*	*S8*	*S9*	*S10*	*S11*
		Seconds						
Non-preferred hand	Precuff	0–5	26	27	27	31	21	27
		5–10	23	24	22	32	24	25
		10–15	22	24	21	30	23	26
		15–20	21	23	18	27	26	21
		Total 0–20	92	92	88	120	94	99
	With cuff	0–5	24	25	34	28	25	24
		5–10	22	22	27	28	27	22
		10–15	21	21	25	26	23	24
		15–20	22	19	27	25	25	23
		Total 0–20	89	87	113	107	100	93
Preferred hand	Precuff	0–5	28	27	30	43	36	29
		5–10	28	27	30	31	32	27
		10–15	31	21	29	33	30	28
		15–20	27	24	26	30	21	30
		Total 0–20	114	99	115	137	119	114
	With cuff	0–5	0	2	16	10	0	0
		5–10	0	6	20	21	3	0
		10–15	4	7	19	12	4	0
		15–20	3	2	21	10	1	0
		Total 0–20	7	17	76	53	8	0

With kinaesthetic and tactile sensation eliminated, i.e. cuff applied for a longer period on the arm of the preferred hand, a marked decrease in tapping is observed.

Here again, as in Table II, all subjects (with the exception of S11) show improvement over the whole or part of the test period.

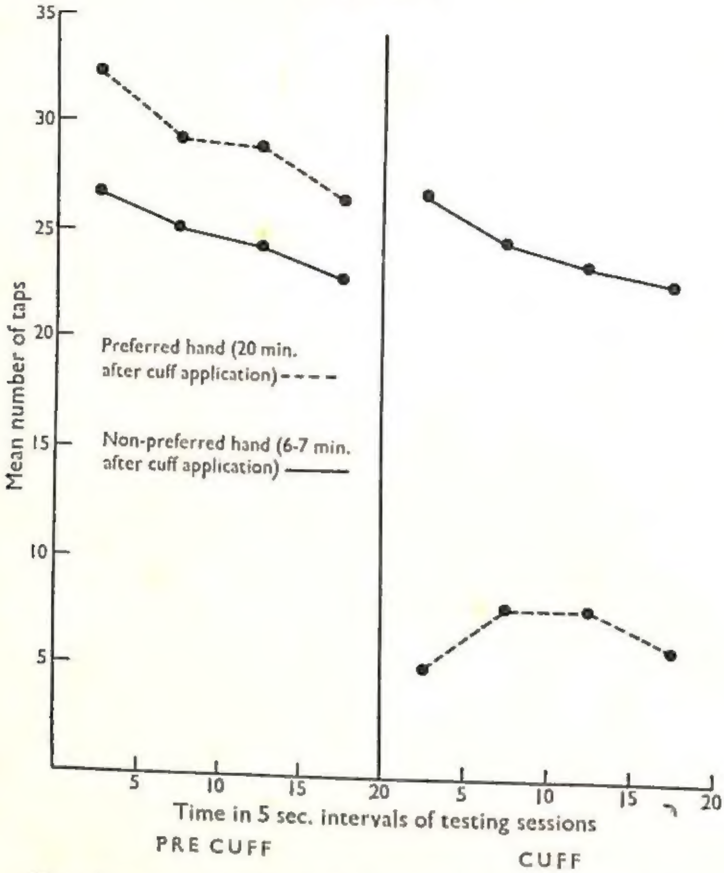
DISCUSSION

The questions posed in the introduction of this paper can now be answered.

Firstly, the mode and time of sense loss can be discussed. The most important sensation—active kinaesthesis—cut-out after 20 min. or more from compression block commencement, in all but one subject (18 min.). The cut-out point was sudden, even dramatic. The subject moved his finger without knowing it. He either omitted to report movement (although asked to do so) or even complained that he was unable to move it, while the finger continued the required movement. When the experimenter assured the subject that he was in effect moving the finger, and asked him to stop, most subjects "refused to believe it," and visually checked. The experimenter herself, being a subject in the experiment, experienced this phenomenon.

Passive kinaesthesis was also eliminated after 20 min. or more but disappeared more gradually. The subjects became unsure of which finger the experimenter

FIGURE 1



Tapping performance of six subjects in Experiment 2.

moved, but reported movement "somewhere" for one or two minutes prior to total sense loss. It must be pointed out, however, that when moving the finger, touch and pressure are involved—thus the finger is subjected to a combination of sensations. It would be impossible, under the present method, to sort out what is actually experienced—pressure, movement, or both.

Tactile sense elimination is the least well defined. As mentioned previously, it tends to cut out then reappear briefly, giving a false cut-off point if not checked further. It is possible that this false cut-off point was due to the random order of testing the fingers. While no statistical check at this stage was available, it seemed that on the ulnar side tactile sensation loss precedes that of the radial side. Final cut-off point generally occurs earlier than the kinaesthetic cut-off point.

This fact is important in so far as it makes removal of kinaesthetic sense without removal of tactile sense impossible with the use of the compression-block method. Whether tactile stimulation is essential in tapping can be determined by using a skin anaesthetic, without interfering with proprioception. Thus its role can be separated experimentally.

The difference in cut-off point for S6 on the two occasions of observation could imply that variations from occasion to occasion of testing for any one individual might vary significantly. An intra-individual reliability test may be of interest to show the extent of variance.

The next problem, that of the relationship between kinaesthetic sense loss and

loss of muscle power, is perhaps the most crucial one when deciding whether compression-block is a useful method in this area or not. If kinaesthetic sense does not cut out before muscle power these two factors would be hopelessly confounded.

Hand squeeze was applied in the present study to clarify this point. This is admittedly a crude method, but it proved rather effective. Tapping a Morse-key is light work compared with a squeeze of the hand. Moreover, hand-squeezing was performed after the tapping task (i.e. O_2 stores were diminished by the tapping).

It could by no means be said that all subjects were able to put normal strength into the hand-squeezing. But only two were incapable of any force, and both these subjects were able to tap. In three cases the subjects reported that they could not do it, while the experimenter had to ask them to stop speedily, as her hand was squeezed to the point of definite discomfort.

The main point here is that muscle power is not affected extensively at the stage where kinaesthetic sensation is already eliminated.

At this point of the discussion of total kinaesthetic sense loss versus partial motor impairment, the following question presents itself: How can we explain that anoxia should affect sensory nerve fibres differently from motor nerve fibres? While the subject reports total loss of movement sensation, he is able to perform the movement exactly as he was asked to. Both motor and proprioceptive fibres belong to the α fibre group which seems to be the type of fibre most sensitive to oxygen lack (DeJong, 1958). However, Sunderland (1951) queries the findings relating to differential sensitivity to ischemia, on the basis of his review of the relevant experimental literature. Neither does the position of the various fibres give a satisfactory explanation. He concludes that, at the present stage of physiological knowledge, no satisfactory explanation is available to account for the finding of differential sensitivity in compression blocks of sensory and motor fibres.

It could be suggested, however, that the difference lies in the sensory receptors, rather than in the differential axon sensitivity. Experimental evidence to that effect is available in work performed on the ear (Johnstone, personal communication) where, in anoxic condition, the receptor activity is the first to be affected, while axonal activity is impaired at a later time only.

Experimental proof of adequate muscle power can be found in the listed results as well. It was pointed out earlier that the majority of subjects improved through the testing session. If failure of muscle power were the factor causing decrement in performance under compression block condition, this could hardly be so. In fact, the opposite results, i.e. diminishing output, would be expected.

It is, however, not possible to go any further than speculation on the above point. Neither, without further experimentation, could any suggestion be made to explain the great individual differences in the extent of performance deterioration here observed. By use of the compression block method, a way may be opened to resolve this question.

Turning now to the remaining questions posed previously; the present experiment seems to show that performance of a skilled motor task, provided it does not involve great muscular exertion, is possible under compression block.

Lastly, it was found that the experimental procedure itself does not interfere with performance. Whether emotional upset, or discomfort, could cause the performance decrement rather than the loss of kinaesthesis was shown to be unlikely. Discomfort is highest in the first 5 min. of the pressure application ("pins-and-needles" are experienced by all subjects). It is presumed that if the subject is worried or even frightened, these reactions would be most pronounced at the beginning of the pressure block. As the experiment progresses, the subject has time to realize

that the procedure is not painful, and at the cessation of the initial "pins-and-needles" the cuff causes no strange sensation. Performance decrement was absent when the subjects were tested at the maximum of emotional upset. Thus, it seems that the ensuing decrement in performance on the preferred arm is due to other than emotional factors.

The above studies show that the loss of kinaesthetic sensation impairs the performance of key-tapping to a marked degree. This pronounced performance decrement gives empirical indication of the importance of kinaesthetic feedback in key-tapping skill. To what degree other tasks would be affected, whether practice could improve performance, and many other related questions could be answered with the help of the compression block technique.

References

Adams, J. A., and Creamer, L. R. (1962). Proprioception variables as determiners of anticipatory timing behaviour. *Human Factors*, **4**, 217–22.

Chase, R. A., Rapin, I., Gilden, L., Sutton, S., and Guilfoyle, J. (1961). Studies on sensory feedback, II. Sensory feedback influences on keytapping motor tasks. *Quart. J. exp. Psychol.*, **13**, 153–67.

DeJong, R. N. (1958). *The Neurologic Examination Incorporating the Fundamentals of Neuroanatomy and Neurophysiology*. New York: Hoeber-Harper.

Provins. K. A. (1958). The effect of peripheral nerve block on the appreciation and execution of finger movements. *J. Physiol.*, **143**, 55–67.

Sinclair, D. C., and Hinshaw, J. R. (1951). Sensory phenomena in experimental nerve block. *Quart. J. exp. Psychol.*, **3**, 49–72.

Sunderland, S. (1951). A classification of periphral nerve injuries producing loss of function. *Brain*, **74**, 491–516.

Manuscript received 27th April, 1965.

VOCALIZATION-AT-PRESENTATION AND IMMEDIATE RECALL, WITH VARYING RECALL METHODS

BY

D. J. MURRAY*

From the Cambridge Psychological Laboratory

In a free recall situation, written recall is superior to spoken recall, and evidence is adduced suggesting that this may be partly due to the fact that this method of recall permits greater freedom in ordering the material. Evidence is also adduced suggesting that voicing the material at presentation gives superior recall because of a facilitatory effect on storage. This interpretation is supported by the finding that the advantage of vocalizing is not affected by prior knowledge as to whether recall is to be oral or written, and by the suggestion that vocalized lists are more resistant to intra-trial interference than non-vocalized lists. In addition, two earlier observations were confirmed: first, that the advantage of voicing is most marked at fast rates of presentation; and secondly, voicing gives rise to a higher proportion of acoustic errors relative to the number of items incorrectly attempted.

Introduction

In previous work (Murray, 1965) it has been shown that if subjects voiced a list of eight consonants, immediate recall of the consonants was superior to that obtained if they whispered them; whispered lists were better recalled than mouthed lists; and mouthed lists were better recalled than silently read lists. The last result only held, however, for lists presented at fast rates (4 letters/sec.). It was also found that written recall was superior to spoken recall, for all vocalization-conditions: and that there appeared to be little interaction between vocalization-conditions and the recall-method.

Before investigating *why* vocalization-at-presentation should lead to improved recall, however, there are a number of findings from the work described above which require rather more detailed examination.

(i) It was found that written recall was superior to spoken ("said") recall overall. Some explanation of this finding is required.

(ii) It seems possible that since subjects knew beforehand whether recall was to be said or written, this information might cause a "set" to try to memorize voiced lists by a particular strategy. To test this, we have compared conditions where subjects knew the recall-method before list presentation, with conditions where subjects were not told of the recall-method until after the list presentation. If voicing-at-presentation preserved its advantage for recall under both conditions, we should be inclined to reject "set" as an explanation of the earlier results.

(iii) There was an indication that voiced lists showed their advantage for recall over non-voiced lists, particularly at fast presentation-rates We need more evidence on this issue.

(iv) An error-analysis revealed that voicing appeared significantly to increase the proportion of acoustic errors to the letters attempted. (By "acoustic errors," we mean recalling "B" as "V," "F" as "X," etc.: cf. Conrad, 1964, for more detailed discussion.) But other types of error, particularly transpositions and serial order intrusions, were not affected by voicing, although their absolute numbers were reduced for voiced lists. These results also require further confirmation.

*Present Address: Department of Psychology, Queen's University, Kingston, Ontario.

The present experiment was therefore devised to resolve some of the above problems. But first, we need to go further into question (i).

In the previous experiment, written recall was almost invariably superior to said recall, for any given vocalization or presentation-rate condition. This is not a result which has always obtained in other memorizing-situations. Koch (1930) found, for instance, that in paired-associate learning, spoken responses were more quickly learnt than were written responses. However, Mackworth (1962) found that written recall was superior to said recall when the stimulus-material consisted of a 3×3 square of digits flashed on to a screen. But it appeared possible that, in our earlier experiment, written recall may have been superior to said recall because subjects could more easily manipulate recall-order while writing. In particular, it might have been easier to write recall in the order last-items-first than to say them in this way. Furthermore, it might have been that retroactive inhibition of items yet to be recalled was higher for said than for written recall. To test this, we have compared written with said recall under a free-recall condition, as used in the earlier experiment, and under a forced-recall condition, in which recall had to be given in the order of presentation, without omissions. If forced recall reduced the difference between said and written recall, we should be inclined to accept the view that freedom of recall-manipulation at least partly determined written recall's advantage over said recall; and by an analysis of serial position effects, we could determine the contribution of intra-recall interference to the observed difference between said and written recall.

Summarizing the above, therefore, we have tried to investigate the reason for written recall's superiority, by comparing "free" with "forced" recall (both said and written) of eight-item consonant-lists; we have tried to control for "set" effects by comparing lists memorized with and without prior knowledge of whether recall would be said or written; and we have tried to repeat the previous experiment more extensively, incorporating the above conditions, in order to check our previous tentative findings that (*a*) voicing did have its advantage most at fast rates, and (*b*) voicing led to a greater proportion of acoustic errors.

Method

General overview. A $2 \times 2 \times 2 \times 2 \times 4$ factorial design was used, each of the 48 subjects doing equal numbers of trials under all conditions. The task consisted of memorizing a list of eight consonants presented visually one letter at a time. The variables were as follows:

(*a*) Voicing vs. non-voicing (reading) at presentation (V vs. R).

(*b*) Said vs. written recall.

(*c*) "Free" vs. "forced" recall.

(*d*) Information as to whether recall would be said or written, given either before ("know before," KB) or after (KA) the presented list.

(*e*) Four rates of presentation: 1, 1·33, 2 and 4 letters/sec.

Details. In many instances, the procedure followed was identical to that described for the earlier experiment (Murray, 1965).

Apparatus. The material to be memorized was stencilled on to a roll of paper, which could be moved past the viewing-slit of a "tracking-machine" (a device similar to a memory-drum). The rate of (continuous) movement of the paper was kept constant but, by differential spacing of the letters on the roll, the various presentation-rates could be achieved.

Material. Lists of eight consonants (e.g. LVMKTNCS) were used. The eight rolls (each consisting of 16 lists) of the earlier experiment were supplemented by eight new rolls: each roll had four lists each at the rates 1, 1·33, 2 and 4 letters/sec. Subsequently,

each roll was cut up into its four groups, and before each experimental session, the four groups were stapled together in a new order. This admitted of a complete ordering of presentation-rate over the whole experiment. (In the earlier study the rates had only been partially ordered over the experiment.)

The lists themselves were devised by a random procedure from the 16 consonants B, C, D, F, H, J, K, L, M, N, P, S, T, V, X and Z. On each *roll* no two lists contained the same letter in the same serial position; within each *group* on the roll, each letter appeared exactly twice.

Presentation. As each list was presented, the subject either silently read or voiced aloud the letters as they appeared. The lists varied in presentation-rate, as described above. But in addition:

(*a*) Subjects were told on half their trials *before* each list whether to "say" or write their recall, and on the remaining trials were told *after* the list whether to say or write their recall. The information was presented via a small illuminated sign situated just above the viewing-slit (see procedure below).

(*b*) Subjects were informed of any change to a new presentation-rate, i.e. after every fourth list. (They had not been informed of this change in the earlier experiment.)

Recall. Written recall was given on slips of paper interleaved with carbon paper. Although a stylus was used, subjects could see what had been written fairly easily. "Said" recall was tape-recorded normally. Instructions for "free" or "forced" recall were given verbally.

Procedure. Written instructions and practice on each condition were first given. The four rolls of that session (see design below) were then run through. On later sessions, the new free/forced and KB/KA conditions were explained verbally to the subject, and a few practice-trials given.

The device used to give the recall-instructions may now be described in more detail. In front of the experimenter was a switchboard connected to a recall-sign box in such a way that, by pre-setting one switch to "KB" and another switch to "say" or "write," the experimenter could turn on a main switch so that: (*a*) the machine would start, (*b*) the appropriate "say" or "write" sign would be illuminated simultaneously, and (*c*) after about 0·75 sec., the light would go out silently and automatically. By this time the list would just be about to appear and the subject would know whether to say or write his recall.

For KA lists, the switch was set to "know after"; in this case the recall-sign would only come on simultaneously with the turning-off of the main switch. Again, it would automatically go out after 0·75 sec. With the "know after" pre-setting, no light would come on when the machine started; with the "know before" pre-setting, no light would come on when the machine was stopped.

At the end of his final session, the subject filled in a written "introspection" questionnaire.

Subjects. Subjects were 48 young adults (mainly from Cambridge University), 33 male and 15 female.

Design. The design of the whole experiment was rather complex, but may be summarized as follows: each of the 48 subjects came for four *sessions* (lasting about 30 min. each). Each session was devoted to four rolls under one of the four conditions KB/free, KB/forced, KA/free and KA/forced. Then within each session two rolls were voiced and two perceived silently; each roll contained 16 lists, four at each presentation-rate; and eight of the lists were recalled by saying, and eight by writing. The "said" and "written" recalls were intermingled over each session to preclude subjects' predicting how the new list would be recalled, but were also arranged so that equal numbers of said and written recall occurred for each presentation-rate and vocalization-condition.

Finally, the following were fully ordered over the whole *experiment*: the four session-conditions (i.e. the KB/KA and free/forced factors); V and R; the rates. As "say" and "write" were fully balanced over the other conditions within each session, we may consider the recall-methods also to be sufficiently balanced across the other variables.

Scoring. As in the earlier experiment the number of letters correct in their original positions was taken as the main index of recall. Also scored were omissions (for "free"

recall only), transpositions, acoustic errors, serial order intrusions, and remainder errors. Detailed definitions of these error-classifications are given in the account of the earlier experiment.

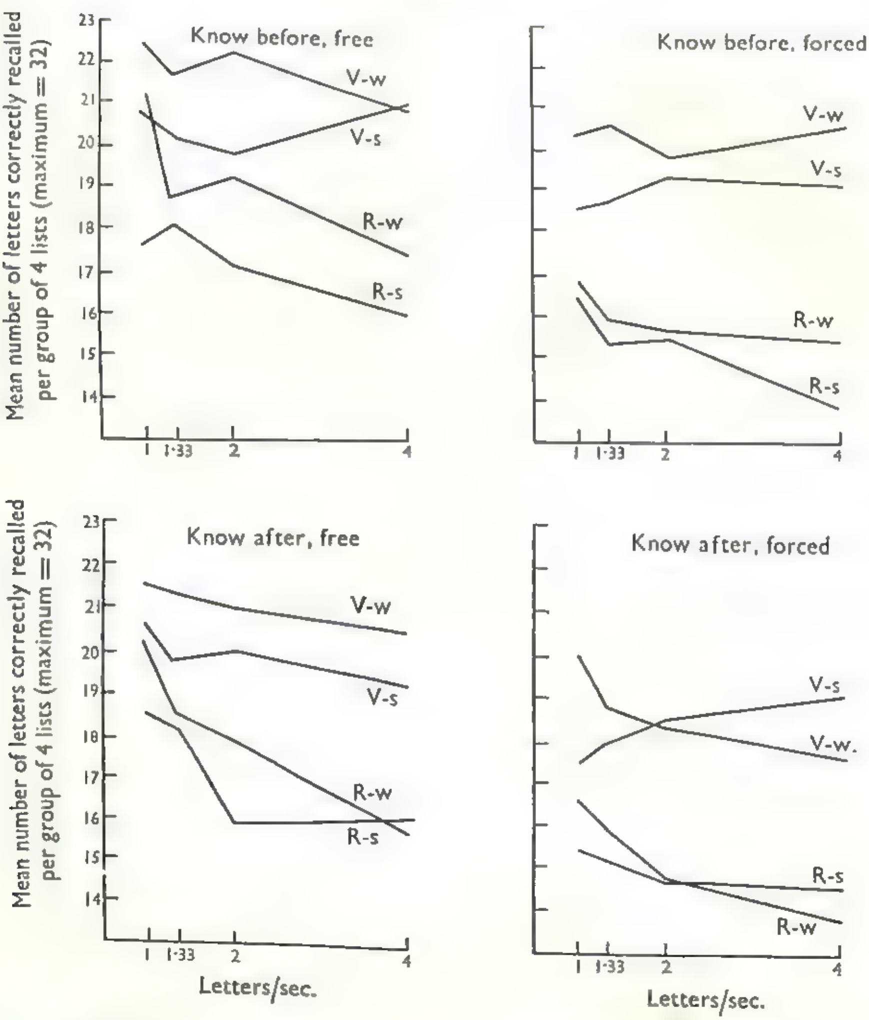
Results

The results will be described under three headings: general results for letters correct; error-analysis; and serial position effects in "forced" recall.

(i) *General results for letters correct*

57·14 per cent. correct recall was obtained overall (in the earlier study it was 57·125 per cent.). Figure 1 shows the overall recall obtained under all conditions. It can be seen that recall of voiced lists (V) was always higher than the recall of

FIGURE 1



Mean number of letters correctly recalled as a function of the conditions of the present experiment.

unvoiced lists (R), at comparable points on the graphs; that forced recall was uniformly lower than free recall; that there was relatively little difference between KB and KA; that in most cases recall was lower for fast than for slow lists; and that in general

written recall was superior to said recall, with the exception of the KA-forced curves. Table I shows extracted means for the say/write variable: the difference between total said and written recall is nearly twice as high on free recall as it is on forced recall. This is reflected in Figure 1 by the fact that the various "said" and written curves on the right-hand graphs (forced recall) appear to be on the whole closer together than the comparable curves on the left-hand graphs (free recall).

TABLE I

EXTRACTED MEANS FOR SAID VERSUS WRITTEN RECALL

	Free recall					*Forced recall*				
	KB		*KA*			*KB*		*KA*		
	R	*V*	*R*	*V*	*Mean*	*R*	*V*	*R*	*V*	*Mean*
Said	17·26	20·43	17·17	19·89	18·68	15·34	18·93	14·96	18·30	16·88
Written	19·17	21·80	18·09	21·05	20·02	16·02	20·32	15·27	18·78	17·59
z	3·96	3·85	2·9	2·95	5·34	0·29	4·12	1·2	1·19	2·68
p	<0·0001	<0·0001	<0·01	<0·001	0·00001	n.s.	<0·0001	n.s.	n.s.	<0·01

Data show mean no. of letters recalled per subject per group of four lists (maximum = 32). On free recall (written - said) = 1·34; on forced recall (written - said) = 0·71.
KB = method of recall known before; *KA* = known after; *R* = read silently; V = voiced.

An analysis of variance was performed on the main results. Table II shows the F ratios obtained for the main factors: significant F-values were found for the factors R/V, free/forced, and subjects, with the say/write factor falling just short of significance at the 0·05 level. Interactions are not shown, as none was found to be significant; however, the highest F-value obtained for any interaction was that between vocalization and presentation-rate (F = 0·64). It is surprising that the say/write value failed to be significant; Table I, however, also shows the results of signs-tests performed on the said vs. written scores. With the important qualification that the results of such tests, applied independently of the general analysis, fail to take account of the various interactions, it appears that, with the exception of KB-V, forcing recall reduced the difference between said and written recall (as was predicted). The interaction say/write × free/forced gave F = 0·44, again one of the higher interactions obtained.

TABLE II

F-VALUES FOR ANALYSES OF VARIANCE ON LETTERS CORRECT AND ERRORS

Variables	*Letters correct*	*Errors (raw scores)*					*Errors (per cent. incorrectly attempted)*			
		O	*T*	*AE*	*SOI*	*Rem.*	*T*	*AE*	*SOI*	*Rem.*
KA, KB	1·71	1·35	0·16	0·03	0·94	0·04	0·009	0·22	0·32	0·07
Free, forced	14·05***	75·6***	17·52***	9·16**	7·20**	15·37***	0·002	0·02	0·003	0·04
R, V	32·32***	8·12**	3·13	0·003	6·84*	11·16***	0·06	3·66†	1·0	2·30†
Say, write	3·23†	0·73	2·49	1·68	0·28	0·05	0·69	0·57	0·22	0·77
Rates (4)	1·08	0·11	0·73	0·16	0·02	0·15	0·22	0·20	0·005	0·02
Subjects (48)	5·88***	1·90***	1·94***	1·03	0·63	1·7***	0·32	0·30	0·009	0·24

*** $p < 0·001$. ** $p < 0·01$. * $p < 0·05$. † p just $> 0·05$.
O = omissions; *T* = transpositions; *AE* = acoustic errors; *SOI* = serial order intrusions; *Rem.* = remainder errors.

It may also appear surprising that the rate factor gave an insignificant variance-ratio. This was confirmed, however, by an application of Jonckheere's test over the rates for the whole experiment (S = 32, $p = 0·07$).

Previously it was found that the difference between V's *advantage* for recall over R for fast rates and that for slow rates was only significant for said and written recall combined. To check this in the present study, we found (V4-R4) and (V1-R1)

for the various recall-methods. The prediction was that, if it were true that voicing showed its advantage most at fast rates, comparisons between these measures should be significant at least for a majority of the comparisons. The results of signs-tests on (V4-R4) vs. (V1-R1) are shown in Table III. It appears that voicing significantly showed its advantage most at fast rates, for one-third of the comparisons. Nevertheless, the signs of the difference all showed (V4-R4) to be greater than (V1-R1).

TABLE III

(V4–R4) versus (V1–R1)—Results of Signs-Tests

	KB-Free	*KB-Forced*	*KA-Free*	*KA-Forced*
Say	$p < 0{\cdot}03$	0·01	n.s.	0·06
Write	0·12	n.s.	0·001	0·16
Combined	0·01	0·06	0·03	0·12

Finally, we originally wished to know whether or not a facilitatory "set" to recall voiced lists by a particular method was operating. A very low F (0·02) was obtained for the interaction between vocalization and say/write. A more detailed examination of the degree by which V exceeded R for said as opposed to written recall showed that for none of the KA/KB or free/forced conditions was V's superiority over R significantly greater on said as opposed to written recall. (The test consisted of finding the percentage by which V exceeded R on each recall condition for each subject, and applying signs-tests to these percentages.)

The KA-forced curves, which appeared somewhat anomalous, are discussed later.

(ii) *Error analysis*

Two types of data will be discussed. (*a*) the raw frequencies of each type of error, and (*b*) the distributions of errors represented as percentages of the number of letters incorrectly attempted, i.e. no. of a particular type × 100 / possible correct, minus actually correct, minus omissions.

Table IV shows both the absolute figures and the error-distributions as percentages of total errors. For the latter, the corresponding data of the earlier experiment are also presented for ease of comparison. It may be seen that the proportions of each error-type are remarkably similar in both experiments. Also shown are the proportions which each error-type formed of the total *possible* in the present experiment: the remaining 57·14 per cent. of the total possible is accounted for by correct recalls.

TABLE IV

Error-Distributions

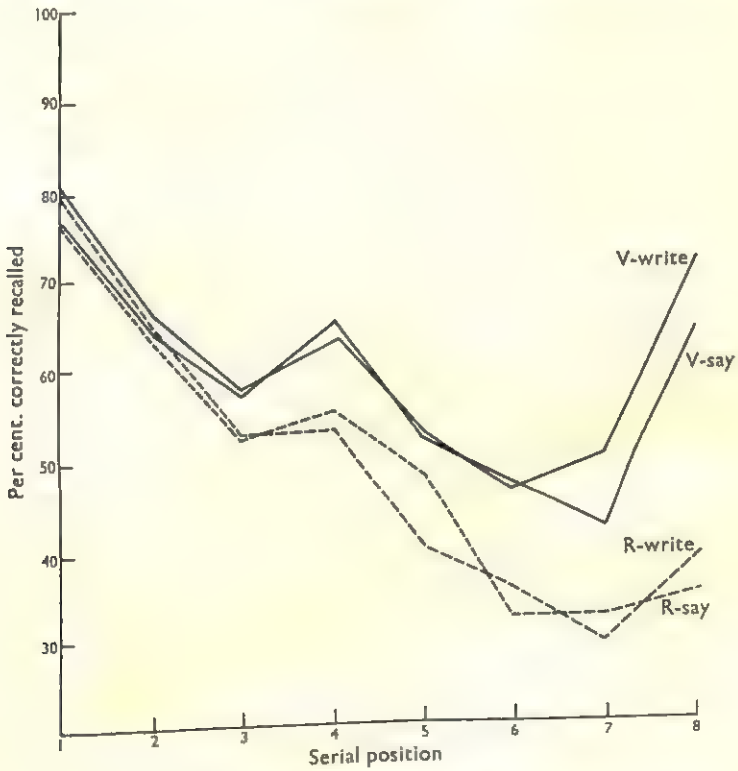
	Omissions	*Trans-position*	*Acoustic errors*	*SOI's*	*Remainder*
Percentage of total possible:					
Free	7·48	12·83	6·67	2·87	9·69
Forced	—	17·88	8·99	4·03	14·75
Percentage of total errors:					
Present experiment ..	17·46	35·82	18·27	8·05	20·4
Previous experiment ..	24·5	30·7	15·7	7·8	21·3

For reasons of space, detailed tables of error-breakdowns under each condition are not given. However, analyses of variance were also performed on the tables of results for each error-type, and the results for the four main error-types are shown in Table II. Again, no interactions were significant, and therefore they are not shown. Remembering that these analyses will not be strictly independent of each other, we may summarize the results as showing that the absolute number of all error-types varied with the free/forced conditions (as might be expected, with no omissions allowed in forced recall); and that some errors varied with R/V, but not acoustic errors. Of the scores expressed as percentages, acoustic errors were nearly significantly higher on V as opposed to R. These results for acoustic errors confirm the results of the earlier experiment.

(iii) *Serial position analysis*

Figure 2 shows the percentage correctly recalled on each position 1 . . . 8 on "forced' recall only. The results over rates and KB/KA conditions are pooled in this graph. The original prediction was that, if saying interfered with recall more than did writing, there should be a greater difference between said and written recall at the later positions of the list than at the early positions of the list. On neither R nor V is that prediction borne out with any conviction.

FIGURE 2



Serial position effects in forced recall, as a function of vocalization and recall method.

However, it is also clear from Figure 2 that V showed its advantage over R particularly on the later positions recalled. This was not a finding which had been predicted. Its importance is discussed below.

An analysis of serial-position effects for the various rates showed that there appeared to be no differential effects of rate on the serial-position curve: all four presentation-rate curves showed the same pattern as those of Figure 2 (a steep fall-off over serial position with upward rises on positions 4 and 8), the curves in general being close together and "parallel."

Summary of results. The main results may be summarized as follows:

(1) V gave significantly superior recall to R: its superiority did not vary with whether recall was said or written: but its superiority was rather more marked at fast rates of presentation than at slow.

(2) Written recall was significantly superior to said recall, on "free" recall; on "forced" recall it was only superior on KB-V.

(3) "Free" recall was overall significantly superior to "forced" recall.

(4) Knowledge of recall-method prior to list-presentation gave significantly superior recall to post-list knowledge of the recall method.

(5) The results of the error-analysis on the whole confirmed those of the earlier study: in particular, voicing increased the *proportion* of acoustic errors relative to other error-types, though not their absolute frequency.

(6) In the serial-position analysis, the curves for "said" and written recall did not diverge at the end of the list to the extent predicted; but the curves for V and R did diverge unduly at the end of the list.

Discussion

We may discuss the above results under the headings given in the introduction.

(i) By "forcing" recall we appear to have reduced by half the advantage of "written" over "said" recall. This suggests that one reason for written recall's superiority over said recall was that it was easier to manipulate written recall during free recall. However, the prediction that written recall should also produce less intra-recall interference was not very well supported.

(ii) The fact that V showed considerable advantage over R both under conditions of pre-list and post-list knowledge of recall-method suggests that V's advantage was not due to a set on the part of the subject to recall by saying or writing particularly. This was further supported by the finding that V's proportional advantage over R was about equal on said and written recall. Again, this suggests that the advantage for recall of vocalized lists lay in their being somehow more efficiently stored than non-vocalized lists.

However, the serial-position analysis throws some new light on the issue. It was found that vocalized lists showed their advantage over non-vocalized lists, particularly on the *last* items recalled. This finding would be consistent with the view that interference during recall was reduced after subjects had voiced the list at presentation, and also with the view that proactive inhibition during presentation was reduced. It is unlikely that a reduction in retroactive inhibition from later members of the presented list was the main factor determined by V, for then the serial position curves should have converged, rather than diverged, at the later positions. It appears that voiced items were stored particularly efficiently, i.e. were resistant to intra-presentation and/or intra-recall interference. Whether RI or PI was dominant is difficult to assess from the present experiment.

But there is, of course, an alternative explanation for the last items being particularly poorly recalled under non-voicing as opposed to voicing: perhaps voiced

lists decayed less rapidly than did non-voiced lists. It is difficult to decide on the present evidence whether a decay-explanation or an interference-explanation is the more valid. But it might be noted that on the decay-hypothesis we should have expected the first items of fast lists to have decayed less than the first items of slow lists: in fact, no differential effect of presentation-rate on the serial-position curves was found. Possibly intra-list rehearsal effectively counteracted the decay on slow lists: and we may remind the reader that V's advantage over R was less marked on slow than on fast lists, in both experiments, a finding consistent with the rehearsal hypothesis. V may also have provided an auditory memory-image for the later items in the list (Mackworth, 1964).

(iii) The present experiment essentially confirmed the finding of the earlier experiment (and also the work of Von Sybel, 1909) that voiced lists showed most advantage for recall at fast presentation-rates. However, it is clear from Table III that the difference is not statistically very reliable. Further discussion of the theoretical importance of this issue will be found in the account of the earlier experiment.

(iv) We now have two experiments in which we have found that the *proportions* of the various error-types to total incorrectly attempted letters were not affected by voicing-at-presentation, with the exceptions of the acoustic errors and the remainder errors. The proportion of the former increased with voicing; the proportion of the latter decreased with voicing. However, in our experiments the acoustic errors have only accounted for about 8 per cent. of the total possible recalled (15 per cent. of total errors). Omissions, transpositions, and remainder errors each account for more, and their overall reduction in absolute number with voicing appears to counteract the slight increase in acoustic errors found on voiced lists, in so far as correct recall is concerned.

But if voicing influences storage, as we have suggested, these other errors may not so much occur during storage as during recall (see Conrad, 1964, for an argument that acoustic errors represent decaying traces). In order to support this claim, we shall briefly report a finding on serial order intrusions. In the earlier experiment we suggested that these might occur if a subject is subjectively distracted during memorization or recall.

In the earlier experiment, subjects were not warned of a change in presentation-rate after every fourth list. In the present experiment, subjects were so warned. It might therefore be predicted, on the above argument, that if the rate changed unexpectedly to a distinctively slower or faster rate, we might expect more serial order intrusions between the two adjacent lists presented at different speeds. In the earlier experiment, frequencies of 58, 35 and 55 serial order intrusions occurred between lists when rates changed between 1 and 1·33, 1·33 and 2, and 2 and 4 letters/sec., respectively. In the present experiment, the equivalent frequencies were 110, 96, and 109. Despite having four times as many lists in the present experiment, only twice as many serial order intrusions of this kind occurred; and for that change of rate which appeared subjectively least surprising, i.e. between the two medium rates, there was a lesser difference relative to the other changes, in the present experiment as opposed to the earlier work (96: 110 or 109, as opposed to 38: 58 or 55). This comparative reduction in frequency of serial order intrusions of this kind, in the present work, suggests that the surprise-factor, reduced in this experiment, was rendered less operative in determining serial order intrusions.

Three other features of the present study may also be noted.

(v) The curves for KA-forced (Fig. 1) differed from most of the other curves in so far as, for both voiced and unvoiced lists, said recall was superior to written recall at the fastest rate (4 letters/sec.). This may have been due to subjects having little opportunity to manipulate written recall in the KA-forced condition.

(vi) The rise in recall on position 4 of the serial position curves was probably due to the adoption by subjects of a grouping technique during memorization, in which they divided the list into two groups of 4 letters (see also Pollack, Johnson, and Knaff, 1959, Expt. III).

(vii) The failure to obtain a significant effect of presentation-rate in the present study may have been related to the possibility that forcing recall appeared to reduce the difference due to presentation-rate (cf. Fig. 1: the "forced" recall curves are in many cases almost horizontal).

In conclusion, then, some of the main results of earlier work have been confirmed; we now have some evidence suggesting that written recall was superior to said recall partly because it allowed more freedom of manipulation of recall-order; and it appears that voicing exerts its effect essentially upon storage, in so far as a "set" to recall by said or written recall did not appear to determine voicing's advantage over non-voicing. Moreover, voiced lists also appeared to be more resistant to interference than did non-voiced lists. It is also suggested that acoustic errors seemed to be more determined by storage-factors than were other error-types.

Some of this work was done while the author was a grateful recipient of a D.S.I.R. grant. He also wishes to thank A. J. Watson for his advice at all stages of the experiment, and Prof. O. L. Zangwill for his continued encouragement. D. Rowell, of the Cambridge Psychological Laboratory, constructed the recall-sign apparatus, and Mrs. M. L. Mutch, of the Cambridge Mathematical Laboratory, programmed the data for computation: their assistance is most gratefully acknowledged.

References

Conrad, R. (1964). Acoustic confusions in immediate memory. *Brit. J. Psychol.*, **55**, 75–84.

Koch, H. L. (1930). Some factors affecting the relative efficiency of certain modes of presenting material for memorizing. *Amer. J. Psychol.*, **42**, 370–88.

Mackworth, J. F. (1962). Presentation-rate and immediate memory. *Canad. J. Psychol.*, **16**, 42–7.

Mackworth, J. F. (1964). Auditory short-term memory. *Canad. J. Psychol.*, **18**, 292–303.

Murray, D. J. (1965). Vocalization-at-presentation and immediate recall, with varying presentation-rates. *Quart. J. exp. Psychol.*, **17**, 47–56.

Pollack, I., Johnson, L. B., and Knaff, P. B. (1959). Running memory span. *J. exp. Psychol.*, **57**, 137–46.

Sybel, A. Von (1909). Über das Zusammenwirken verschiedener Sinnesgebiete bei Gedächtnisleistungen. *Z. Psychol.*, **53**, 257–360.

Manuscript received 21st April, 1965.

THE ROLE OF ATTENTION IN THE DISINHIBITION OF DISPLACEMENT ACTIVITIES

BY

D. J. McFARLAND

From the Department of Psychology, University of Durham

A mechanism for the disinhibition of displacement activities is proposed, by which a factor arising from non-reward causes attention to be switched to irrelevant aspects of the stimulus situation. In addition to evidence already extant, three lines of evidence are presented for the case of displacement pecking in the Barbary dove: (1) partially rewarded doves learn more about incidental aspects of a discrimination situation than consistently rewarded doves; (2) partially rewarded doves respond more readily to irrelevant stimuli than consistently rewarded doves; (3) doves which take longest to extinguish a simple approach response, also spend most time displacement pecking.

Introduction

The detailed causal analysis of displacement activities in recent years has brought into focus the importance of competition between different activities. The historical background to this development has been admirably reviewed by Zeigler (1964) and need not be discussed here, except to say that a displacement activity is an activity which belongs to a motivational system other than those prominently activated.

Andrew (1956) suggested that displacement grooming in buntings is controlled directly by the peripheral stimuli which normally elicit the behaviour, and indirectly by the strength of other activities, incompatible with grooming. He showed that grooming is easily suppressed by other activities and postulated that the peripheral stimuli were continually present. In conflict situations, when neither of the primary competing activities can be fully expressed, the suppression of grooming is sufficiently reduced to allow "displacement" grooming to occur. This "disinhibition" theory of displacement was stated more explicitly by van Iersel and Bol (1958), who postulated that equilibrium between two conflicting "drives" lowered their power to inhibit a third "drive," which could then be expressed as overt behaviour. The work of Sevenster (1961) and of Rowell (1961) reinforces the disinhibition hypothesis by confirming the evidence upon which it is based. This evidence may be summarized as follows:

(1) In conflict situations, displacement activities tend to occur when the conflicting activities are near equilibrium, so that neither can occur.

(2) Displacement activities can be influenced both quantitatively and qualitatively by external factors which normally influence these activities.

(3) Displacement activities are usually prepotent in the animal's repertoire in the sense that they are generally common activities, such as grooming and feeding, or are seasonally attuned to the animal's motivational state, such as reproductive activities (Bindra, 1959).

Although the disinhibition hypothesis seems satisfactory as a general explanation for the occurrence of displacement activities, there are still a number of questions to be answered.

Reviewers of field studies (e.g. Tinbergen, 1952; Bastock, Morris and Moynihan, 1953) have noted that displacement activities tend to occur in three types of situation:

(1) physical thwarting of appetitive behaviour, (2) thwarting of consummatory behaviour by removal of the objective, (3) simultaneous activation of incompatible behaviour patterns. These observations have, to some extent, been confirmed in the laboratory (McFarland, 1965). The disinhibition theorists to date envisage a mechanism for disinhibition based on a reciprocal inhibition between two "behavioural systems" (Sevenster, 1961). However, displacement activities have been shown to occur in two of the types of situation mentioned above, in which it is difficult to see how there could be conflict between two behavioural systems, since there would seem to be only one major motivational system involved. There may, of course, be weak "secondary drives" induced by frustration, but to invoke these introduces a different type of theory from the disinhibition theory, which requires two systems of equivalent strength to be simultaneously active.

Another difficulty arises from the use of blanket terms such as "drive" and "tendency." Precisely what is inhibited and disinhibited? This question cannot be answered until the mechanism of disinhibition has been worked out. The aim of this paper is to contribute to this task.

A possible mechanism of disinhibition

Activities are said to be incompatible when they cannot occur simultaneously. Sherrington (1906) pointed out that activities which share motor components cannot occur simultaneously, and must therefore compete for control of the motor mechanism. He introduced the principle of reciprocal inhibition as a vehicle for such response competition. Extension of this principle to all types of competing activity would mean that there must be multiple reciprocal inhibition between all the potential activities that are capable of being expressed in any particular situation. In such a case only the activity with the strongest potential would gain control of the motor mechanism and could be observed. However, the disinhibition hypothesis implies that a displacement activity is not an activity which has the strongest potential at the time of observation, but an activity which occurs when other activities with stronger potentials are blocked. The question is, how does this blockage allow an inferior potential activity to become overt? One possibility emerges on consideration of stimulus competition as an alternative or additional mechanism producing incompatability. If an animal paid attention to only a few aspects of the stimulus situation within any one time period, the number of activities which could be "released" during that time, would be limited. Clearly, the extent to which the animal switches attention from one stimulus to another must play a large part in determining which of the potential activities becomes dominant. It is here suggested that, *when an ongoing activity is blocked, attention is switched to stimuli other than those eliciting the ongoing activity*. An activity can be blocked in any of the three ways mentioned above, and displacement occurs in all three situations by being disinhibited via a switch of attention.

Evidence for the existence of an attention mechanism in animals comes from a number of different types of study. There is a considerable body of work which indicates that humans have a limited capacity to handle the information provided by their sense organs, and that there is a central selecting mechanism, which allows some of this information to be utilised in determining behaviour, and excludes the rest. Much of this evidence, which has been reviewed by Broadbent (1958), comes from experiments concerned with the subject's ability to handle information presented simultaneously to different ears, or to ear and eye. Physiological evidence for central control of sensory inflow to the brain has implicated the reticular activating system (Hernandez-Peon, 1961) and a number of physiological manifestations of the attentive

process have been claimed. There is, however, little agreement, amongst physiologists, as to the mechanisms involved. Considerable evidence for attention-like processes comes from behavioural work on animals, and this work is directly relevant to the present argument.

When an animal is required to discriminate between two stimuli in order to obtain a reward, there are often a number of cues, such as shape, size, brightness, etc., which it can utilise. Bruner, Matter and Papanek (1955) showed that the more rats were trained with only one cue relevant to the solution of a discrimination, the less they learned about a second cue, which was additionally present during the latter part of training. A number of workers (e.g. Lawrence, 1950; Pubols, 1956; Mackintosh, 1962) have shown that the more training given on a certain discrimination, the more easily animals learn to reverse that discrimination. Sutherland (1959) pointed out that such apparently paradoxical results can be explained on the assumption that discrimination learning is a two-stage process. He suggested that animals learn (1) which analysing mechanism to switch in (which type of cue to attend to), and (2) which response to attach to the output of the relevant analyser. Reward increases the probability that the analyser used on that trial will be used in the future. Thus the more an animal is trained with a particular cue relevant, the more firmly will the corresponding analyser become established for use in that situation. Overtrained animals will, therefore, have the relevant analyser more firmly established than non-overtrained animals, and will consequently be better equipped for reversal. This theory implies that non-overtrained animals will have a greater tendency to attend to incidental cues during reversal, and Mackintosh (1963) has shown that this is in fact the case. Sutherland (1964) extended this implication to predict that partially rewarded (PR) animals learn about a greater number of incidental cues than consistently rewarded (CR) animals. During extinction each animal has to extinguish the response to all the analysers established during training. Thus those with more analysers switched in should take longer to extinguish. In this way Sutherland is able to explain the well known, but ill explained, fact that PR animals take longer to extinguish than CR animals.

When animals are reversed in a discrimination experiment, they continue to respond to the stimulus which was previously correct and consequently are frequently non-rewarded. Mackintosh showed that, when irrelevant cues were present during reversal, the relearning of non-overtrained rats was retarded much more than that of overtrained rats. However, it made little difference whether or not irrelevant cues were additionally present prior to reversal. According to Sutherland's theory, the greater effect of irrelevant cues on non-overtrained rats is due to the relevant analyser being less well established. But it is clear from Mackintosh's experiment that this only becomes important when irrelevant cues are present during reversal. Is it possible that, *as a result of being non-rewarded* during reversal, animals switch attention to irrelevant cues, unless they already have an analyser very firmly established? Similarly, if PR animals learn more about incidental cues, is it possible that non-reward is again responsible? Non-reward of this type corresponds to one of the three "blocking" situations in which displacement activities have been shown to occur. Thus it is possible that the mechanism which operates in the displacement situation also lies behind Sutherland's theory.

EXPERIMENT I

This experiment was designed to determine whether PR animals do in fact learn more about incidental cues than CR animals.

Method

The subjects were sixteen experimentally naïve adult Barbary doves (*Streptopelia risoria*), which had been bred in the colony at the Institute of Experimental Psychology, Oxford.

The doves were trained and tested in a discrimination box, in which the starting box measured 16 × 20 cm. and was fitted with a transparent perspex detention door. This door opened at the apex of a triangular-shaped chamber, the base of which was formed by two 20 × 20 cm. doors, hinged at the top and separated by a vertical partition projecting 8 cm. towards the starting box. The apparatus was painted grey throughout the inside. The discriminanda were cut from black or white ⅛-in. perspex sheet, and were fastened, one to each of the two doors opposite the starting box.

A dove was released from the starting box at the beginning of each trial and was kept in the apparatus until it made a choice. The criterion of choice was met when the bird placed both feet across the line running parallel to the stimulus doors and flush with the end of the partition. At each choice the corresponding door was opened by the experimenter, revealing a food dish which was either filled with grain (+) or was empty (−). In both cases the bird was allowed 10 sec. access to the food dish before being removed from the apparatus. After each day's testing, the doves were allowed 2 min. access to food in the home cage, after which food was withheld until the next day. Water was always available in the home cages.

Pretraining

On the first day of the experiment each bird was left in the apparatus for 2 hr. with access to both food dishes, but with no stimuli on the doors. On the second day he was given five trials under the same conditions. On the third day the treatment was the same except that the doors were closed at the outset of each trial. Training proper was begun on the fourth day.

Training

The birds received ten trials per day for the rest of the experiment. They were trained, first, to discriminate between black and white 7 cm. squares. For half of them only black choices were rewarded, for the other half, white. Half of those choosing each stimulus were rewarded for every correct choice (CR group), while the other half were rewarded for only 50 per cent. of their correct choices (PR group). The position of the positive and negative stimuli and the order of the rewarded trials was varied according to selected Gellermann orders. As each bird reached a criterion of eight correct choices on three successive days, it was transferred to the second problem.

With respect to brightness and to reinforcement schedule, the second problem was the same as the first, but the stimuli were changed to 4 × 12 cm. rectangles, so that the problem could also be solved in terms of orientation. For two birds in each subgroup, horizontal was positive; for the other two, vertical was positive. Each subject was given 60 trials on this problem.

Testing

The subjects were tested to determine (1) what had been learned about orientation, and (2) the resistance to extinction. Each bird received 20 non-rewarded transfer tests, each transfer test alternating with a retraining trial on the second problem (rectangles). During the transfer tests the discriminanda were either white horizontal versus white vertical, or black horizontal versus black vertical. The transfer tests were followed by 20 further retraining trials and then by 20 extinction trials, also using the rectangular stimuli. During the extinction trials the time taken to make a choice was measured with a stopwatch.

Results

At the end of the first phase of the experiment, the PR group had taken a mean of 90·8 trials to reach criterion and the CR group averaged 87·5 trials. On the last day of phase one, the PR group obtained an average of 8·7 correct choices and the CR group averaged 8·8 correct choices. These results show that the learning speed of the two groups was equivalent during the first phase of the experiment.

The results of the transfer tests are presented in two ways. Firstly, the absolute number of correct choices made by the PR group is compared with that made by the CR group. Thus the former obtained a mean score of 15·0 out of 20, while the latter obtained only 10·4 out of 20 trials. By the Mann-Whitney U test (one-tailed) $U = 4{\cdot}5$, $p < 0{\cdot}005$ (two of the PR scores were discounted because the birds refused to choose). A more sensitive test is obtained by subtracting the score obtained by each bird on the transfer trials from that obtained during the alternating retraining trials. The mean difference for the PR group was 2·0, while that for the CR group was 8·75 (Mann-Whitney $U = 1{\cdot}5$, $p = 0{\cdot}001$). Clearly the PR group did much better on the transfer tests than the CR group. Thus the prediction made by Sutherland (1964) is confirmed. In order to determine whether Sutherland's theory about the effect of PR on extinction could also be supported, the birds were tested to determine whether they had received sufficient PR to affect extinction. Over the first 10 extinction trials there was little difference in choice time between the PR and CR groups, but over the last 10 trials the mean score for the CR group was 46·8 sec., while that for the PR group was only 12·2 sec. (Mann-Whitney $U = 0$, $p \ll 0{\cdot}001$). Clearly, the CR group were extinguishing faster than the PR group. Now if Sutherland's theory is correct, those subjects, within each group, which learn most about the incidental cue, should also show the greatest resistance to extinction. Ranking the subjects within each group by (1) the magnitude of the difference between the scores on the transfer and retraining trials, and (2) the mean choice time over the last 10 extinction trials, the Kendall Rank correlation test indicated that there was no significant association between the two variables for the PR group ($S = 5$, $p = 0{\cdot}023$), but for the CR group there was a significant association ($S = 15$, $p < 0{\cdot}015$). The failure to find any correlation within the PR group may well be due to the fact that they had not yet begun to extinguish. Thus the mean time difference between the first 10 and the last 10 extinction trials was 34·0 sec for the CR group and only 2·7 sec. for the PR group.

It is concluded that the prediction that PR animals learn more about incidental cues is confirmed by this experiment, but that the question of extinction requires further investigation.

Experiment 2

The results of Experiment 1 show that partially rewarded Barbary doves learn more about incidental aspects of the test situation than do consistently rewarded animals. Sutherland (1965) has obtained similar results from rats. The implication is that PR animals are more likely to attend to incidental stimuli on any particular trial. If this is so, then PR animals should also be more likely to respond to stimuli newly introduced into the test situation. This prediction is tested in this expeiment.

Method

Fifteen experimentally naïve adult Barbary doves were used. They were caged socially at controlled temperature (60°–70° F.) and artificial daylength (8 hr. light: 16 hr. dark).

Pretraining

During pretraining the birds were deprived of water in the home cage. Water was available only in the testing apparatus, which consisted of a 60 × 18 × 18 in. enclosed runway, painted black inside. They entered the runway from a starting box at one end, and were trained to run to the other end, where they were allowed to drink their fill. Each bird was given one pretraining trial per day for 10 days.

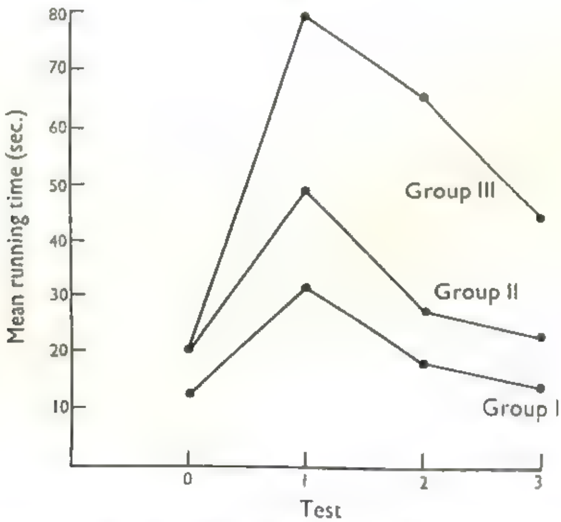
Training

At the end of pretraining the doves were divided into three equal groups, matched for mean running speed. Each group received two training trials per day for 15 days. Groups I and II were rewarded with 3 c.c. of water on each trial and were deprived of water in the home cage. Group III were rewarded on only one trial each day, reward being randomly alternated with non-reward. Each bird in group III was given a further 3 c.c. of water in the home cage at the end of each training session.

Testing

At the end of training the birds were allowed to recover from deprivation, being given *ad lib.* food and water in the home cage, for 3 days. Group I was then deprived of water for 48 hr. and groups II and III for only 24 hr. After deprivation each bird was given a trial in the runway, the time from entering the runway to drinking being recorded. This preliminary test (test O) was identical to the training trials and was followed, after recovery and redeprivation, by three testing trials proper. The procedure for these tests was identical to that of the preliminary test, except that a red flashing light was present above the drinking trough at the far end of the runway, and visible to any bird entering the runway. The light was switched on as soon as the bird had entered the runway. The rationale of this testing procedure was as follows.

FIGURE 1



The effect of a deterrent on the running time of three groups of birds. Group I was CR and 48 hr. thirsty; group II, CR and 24 hr. thirsty; group III, PR and 24 hr. thirsty.

According to the present theory, partially rewarded animals should respond more to novel stimuli, because they have a greater tendency to attend to incidental aspects of the situation. Thus group III should be affected more by the red light than groups I and II. Because Barbary doves are normally frightened by novel stimuli, it is expected that the effect of the red light will be to lengthen the response time, but because group I are more thirsty than group II, this effect is expected to be smaller in group I.

Results

From Figure 1 it can be seen that the effect of the flashing light was to increase the response time for all the experimental groups. With each successive test the response time decreased, indicating a general habituation effect. The response time for group I was consistently less than that for group II (tests 0, 1 and 3, $t > 1{\cdot}91$, one-tailed, $p < 0{\cdot}05$), indicating that increased deprivation contributes to the approach tendency. Comparing the performance of groups II and III, using one-tailed tests; for test 1, $t = 2{\cdot}31$ ($p < 0{\cdot}025$) and for test 2, $t = 2{\cdot}48$ ($p < 0{\cdot}025$),

showing that the different training procedures to which groups II and III were subjected made an important difference to their test performance.

Discussion

The results of this experiment support the view that PR animals are more likely to attend to stimuli newly introduced into the test situation. It is interesting to compare the results of this experiment with those which would be predicted from Lawrence and Festinger's (1962) theory of cognitive dissonance.

In applying the theory of cognitive dissonance to animals, Lawrence and Festinger maintain that a rat which experiences non-reward develops *dissonance*, a tensional state which the animal tries to reduce. Dissonance reduction can be brought about either by changes in behaviour, or by finding *extra attractions* in the situation in which dissonance was experienced. During a PR schedule the extra attractions generated lead to an accumulation of *dissonance motivation*. This extra motivation causes rats to persist longer in extinction, thus producing the characteristic partial reward effect on extinction. If PR produces dissonance motivation, then clearly, in the present experiment both the PR group (group III) and the more thirsty group (group I) should be less affected by the red light than group II; in contrast to the prediction derived from Sutherland's theory, which is that the PR group will be more affected than the controls. It appears from Figure 1 that Sutherland's theory offers the better interpretation of the results of this experiment.

However, these results can only be taken as criticism of cognitive dissonance if group III was in fact more resistant to extinction than group II. To test this proviso, groups II and III were tested for resistance to extinction at the end of the experiment. The extinction trials were run all on one day. Each dove had about 20 min. between trials and was run until it reached a criterion of three accumulated trials with latencies greater than 180 sec. The number of trials taken to reach criterion was the measure for resistance to extinction. This measure was found to be 16·8 for group II and 22·8 for group III ($t = 3{\cdot}2$, $p < 0{\cdot}01$). Thus, although the testing procedure had a greater deterrent effect on group III, this group was more resistant to extinction than group II. Thus it is possible that the degree to which animals switch attention governs both their responsiveness to incidental stimuli and their resistance to extinction.

The relationship between displacement and diversion of attention

The experiments reported so far provide evidence for the view that Barbary doves, which are non-rewarded, (1) learn more about irrelevant cues, and (2) show greater response to novel stimuli. These results support the present interpretation of Sutherland's theory, so that, in all, there is considerable evidence for the view that non-reward switches attention. In the introduction to this paper it was pointed out that such a mechanism could be important in the disinhibition of displacement activities. Thwarting of both appetitive and of consummatory behaviour represent non-reward situations, and both are inductive to displacement. Displacement also occurs in conflict situations, which also provide conditions in which reward is blocked or delayed. However, the role of attention is likely to be more complicated in a conflict situation, and it is necessary, first of all, to determine whether there is any relationship between displacement and diversion of attention in a simple non-reward situation.

As attention cannot be measured directly, it is necessary to estimate it through one of the behaviour variables which are supposed to correlate with attention. The amount learned about incidental cues, the degree of response to incidental stimuli,

and the resistance to extinction, are all, on the present interpretation of Sutherland's theory, variables of this kind. One way to tackle the problem is to determine whether individual differences for these variables correlate with the amount of displacement activity shown by the same individuals in a test situation. This rationale forms the basis of the next experiment.

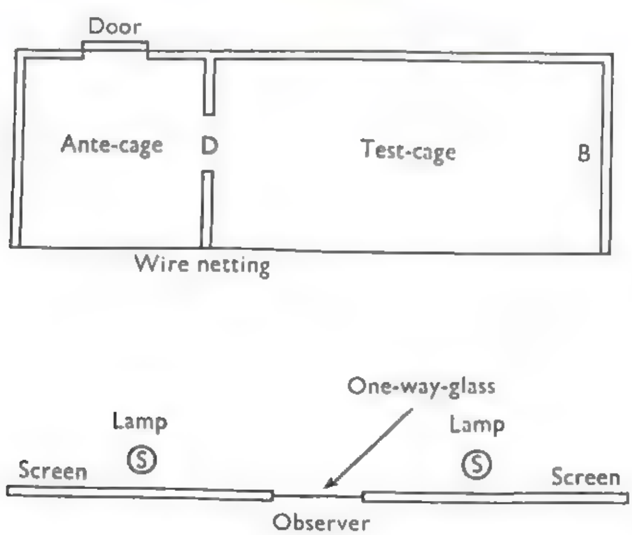
Experiment 3

Because an individual's displacement activity varies greatly from trial to trial, it is necessary to devise a situation in which the mean displacement activity can be obtained from a number of trials. From this point of view, the extinction situation is the most suitable of the attention-measuring situations mentioned above.

Method

Ten adult Barbary doves were trained in the apparatus shown in Figure 2. They entered the antecage individually and ran to the water bowl (B) when the door (D) was opened by the experimenter. Each dove was given one trial a day for 20 days and was allowed to drink its fill on reaching the water bowl. No water was available in the home cage, but food was available *ad lib.* At the end of training all birds were reaching the water bowl within 30 sec. of the door (D) being opened, and all were then transferred to an extinction schedule. During the extinction tests each bird continued to be run once per day. The water bowl in the test-cage was empty, and each bird was allowed one drink in the home cage at the end of the daily testing session. Each extinction trial lasted 180 sec. at the end of which the bird was removed from the apparatus, and its identity noted from the number on its leg ring. During extinction the time taken to reach the water bowl (response time) was measured with a stopwatch. If a dove failed to touch the water bowl within 180 sec., on three separate occasions, it had reached the criterion of extinction. As each dove reached this criterion it was removed from the experiment. Throughout extinction the time spent pecking at the ground, both before and after the bird reached the water bowl, was recorded on electromagnetic cumulative counters.

FIGURE 2



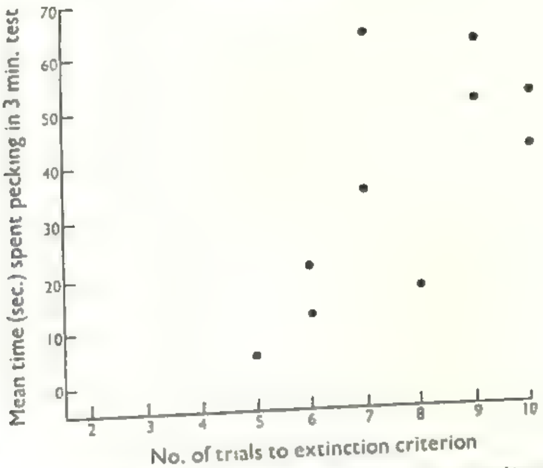
Apparatus used in Experiment 3. D = door, remotely controlled by the observer; B = water bowl.

Results and discussion

The results of this experiment are summarized in Figures 3 and 4. From these results it can be seen that there is a positive correlation between the mean time

spent pecking during the extinction tests and the number of trials taken to reach the extinction criterion (Spearman rank correlation coefficient $r_s = 0{\cdot}65$, $t = 2{\cdot}40$, $p < 0{\cdot}025$). There is also a correlation between pecking and the mean response time ($r_s = 0{\cdot}564$, $p < 0{\cdot}05$). In other words, those birds which were most resistant to extinction also showed most displacement activity.

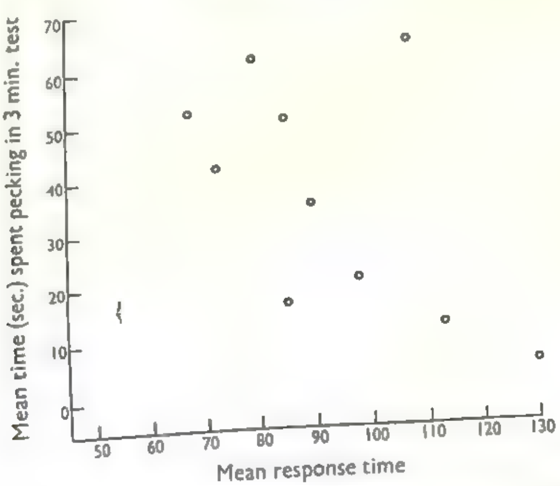
FIGURE 3



Correlation between resistance to extinction (trials to criterion) and time spent displacement pecking during extinction.

It should be noted here that the usual procedure of removing the bird from the apparatus when it reached the goal was not followed. In the present experiment there was no correlation between the amount of pecking before the goal was reached and the resistance to extinction. The reason for this is simply that the birds most resistant to extinction run faster and therefore have less opportunity for displacement. Only by leaving the bird in the apparatus for a fixed amount of time can a true measure of its displacement activity be obtained.

FIGURE 4



Correlation between resistance to extinction (mean response time) and time spent displacement pecking during extinction.

According to the present theory, resistance to extinction is a function of the extent to which the birds pay attention to incidental cues during training. The finding that the birds more resistant to extinction do more displacement pecking supports the view that displacement activities are dependent on some attention switching process. There are two possible ways in which this process may affect displacement during extinction. First, having many analysers switched in (Sutherland's terminology) causes the subjects to be more resistant to extinction, because they have to learn to extinguish the responses to these analysers; in addition, it causes them to spend more time displacement pecking, because they have learned to take notice of incidental stimuli, as suggested by the results of Expt. 2. The alternative is that those subjects which have a large number of analysers switched in do more displacement, and *consequently* devote less time to learning about the relevant stimuli, so that they extinguish slower.

In order to distinguish between these two possibilities, the doves were retrained and then divided into two equal groups, matched for previous resistance to extinction. During retraining both groups experienced the test-cage with the floor covered with small particles of sand, faeces, etc.; and with the floor swept clean. One group was then extinguished with grain scattered over the floor and the other with the floor swept clean. This procedure is based on previous experiments (McFarland, 1965), which showed that the presence of grain greatly facilitates displacement pecking. Thus, if the second alternative is correct, the birds with grain should spend more time pecking and should consequently take longer to extinguish. If the first alternative is correct, the induced displacement pecking should make no difference to the resistance to extinction. The results of the extinction tests show that the presence of grain made a large difference to the time spent pecking (grain group mean = 55·1 sec./trial; no-grain group mean = 21·9 sec./trial), but there was no difference in the resistance to extinction (grain group mean = 8·0 trials to criterion; no-grain group mean = 9·0). These results suggest that displacement is independent of resistance to extinction, but shares with it a common causal factor.

Discussion

Lambert and Solomon (1952) trained rats in a runway and then physically blocked them at different distances from the goal. Rats blocked near the goal were found to be more resistant to extinction and also show more "excitement," including irrelevant activity. Clearly Lambert and Solomon obtained essentially the same relationship between displacement activity and resistance to extinction as that reported in this paper. If these findings were to be substantiated they could have important implications for extinction theory. In particular, frustration theory (e.g. Amsel, 1958) maintains that PR animals are more resistant to extinction because they have become used to the disrupting effects of frustration generated by non-reward. It has also been argued that since frustration has drive properties, PR animals, which experience frustration during training, should run faster than CR animals. This appears to be the case for the later part of acquisition (Weinstock, 1958; Goodrich, 1959), but it has been found that the response strength of PR animals is weaker during the early part. According to Weinstock and Goodrich, frustration drive energises competing responses, which detract from the overall response measure (Cotton, 1953). Later in training the competing responses extinguish and the frustration drive can only energise the relevant response. On this basis one would expect PR animals to show less competing (displacement) responses during extinction, partly because PR animals are less frustrated than CR animals, and partly because

the competing responses are supposed to have been extinguished. On the present hypothesis, of course, PR animals should show more competing responses during extinction. Clearly the relationship between displacement and extinction has important theoretical implications and is likely to be a fruitful field for research.

The other main implication of the relationship between displacement and extinction reported above, is that it suggests that animals which attend to incidental stimuli also spend more time in displacement activity. McFarland (1966) has data which suggest that this may also occur in conflict and in thwarting situations. In an approach-avoidance conflict situation McFarland obtained a positive correlation between pecking and a stationary "attentive" posture (SAP) and a negative correlation with a stationary ambivalent posture (SAV). It is argued that an animal in SAV is in no position to attend to irrelevant aspects of the situation because the very nature of the posture (approach-avoidance compromise) implies that attention is fixed on relevant stimuli. When Barbary doves are physically prevented from reaching a goal, they do less SAV but more SAP and pecking than they do when in conflict. These data support the view that displacement activity and attentiveness are connected.

The thesis presented here is that "frustration," which may result from conflict, thwarting or non-reward, diverts "attention." This mechanism is not the sole cause of displacement; on the contrary, it is evident from a number of experiments (e.g. Andrew, 1956; Rowell, 1961; McFarland, 1965) that there must also be an underlying motivational readiness to perform the type of activity which appears as displacement. Disinhibition theory rejects the view that displacement activities result from displaced motivation and maintains that they are energised by their normal motivational factors. It has been argued that the term "displacement" should be dropped, although some (e.g. Sevenster, 1961) prefer to retain it for historical reasons. However, in this paper the view is put forward that in order for an animal to respond to underlying motivations, its attention must be displaced from the stimuli relevant to the predominant motivation. It is hoped that the term "displacement" will thus acquire a new lease of life.

Two of the experiments reported in this paper were conducted at the Institute of Experimental Psychology, Oxford. The writer wishes to express his appreciation to J. D. Delius, B. McGonigle, N. J. Mackintosh and D. M. Vowles for reading the manuscript and for their helpful comments.

References

Amsel, A. (1958). The role of frustrative nonreward in noncontinuous reward situations. *Psychol. Bull.*, **55**, 102–19.

Andrew, R. J. (1956). Normal and irrelevant toilet behaviour in *Emberiza* spp. *Brit. J. anim. Behav.*, **4**, 85–91.

Bastock, M., Morris, D., and Moynihan, M. (1953). Some comments on conflict and thwarting in animals. *Behaviour*, **6**, 66–84.

Bindra, D. (1959). An interpretation of the 'displacement phenomenon'. *Brit. J. Psychol.*, **56**, 263–68.

Broadbent, D. E. (1958). *Perception and Communication*. London: Pergamon.

Bruner, J., Matter, J., and Papanek, M. (1955). Breadth of learning as a function of drive level and mechanization. *Psychol. Rev.*, **62**, 1–10.

Cotton, J. W. (1953). Running time as a function of amount of food deprivation. *J. exp. Psychol.*, **46**, 188–98.

Goodrich, K. P. (1959). Performance in different segments of an instrumental response chain as a function of reinforcement schedule. *J. exp. Psychol.*, **57**, 57–63.

Hernandez-Peon, R. (1961). Reticular mechanisms of sensory control. In Rosenblith, W. A. (Ed.), *Sensory Communication*. New York: M.I.T. Press.

IERSEL, J. A. A. VAN, and BOL, A. C. A. (1958). Preening in two tern species. A study of displacement activities. *Behaviour*, **13**, 1–88.

LAMBERT, W. W., and SOLOMON, R. L. (1952). Extinction of a running response as a function of distance of block point from the goal. *J. comp. physiol. Psychol.*, **45**, 269–79.

LAWRENCE, D. H. (1950). Acquired distinctiveness of cues: II. Selective association in a constant stimulus situation. *J. exp. Psychol.*, **40**, 175–88.

LAWRENCE, D. H., and FESTINGER, L. (1962). *Deterrents and Reinforcement*. London: Tavistock.

McFARLAND, D. J. (1965). Hunger, thirst and displacement pecking in the Barbary dove. *Anim. Behav*, 293–300.

McFARLAND, D. J. (1966). On the causal and functional significance of displacement activities. *Z. Tierpsychol* (in press).

MACKINTOSH, N. J. (1962). The effects of overtraining on a reversal and a nonreversal shift. *J. comp. physiol. Psychol.*, **55**, 555–9.

MACKINTOSH, N. J. (1963). The effect of irrelevant cues on reversal learning in the rat. *Brit. J. Psychol.*, **54**, 127–34.

PUBOLS, B. H., Jr. (1956). The facilitation of visual and spatial discrimination reversal by overlearning. *J. comp. physiol. Psychol.*, **49**, 243–8.

ROWELL, C. H. F. (1961). Displacement grooming in the chaffinch. *Anim. Behav.*, **9**, 38–63.

SEVENSTER, P. A. (1961). A causal analysis of a displacement activity (fanning in *Gasteroteus Aculeatus* L.). *Behaviour*, Suppl. No. 9.

SHERRINGTON, C. S. (1906). *The Integrative Action of the Nervous System*. Yale: University Press.

SUTHERLAND, N. S. (1959). Stimulus analysing mechanisms. In *Proceedings of a Symposium on the Mechanisation of Thought Processes*. Vol. 2. London. H.M.S.O. Pp. 575–609.

SUTHERLAND, N. S. (1964). Visual discrimination in animals. *Brit. Med. Bull.*, **20**, 54–9.

SUTHERLAND, N. S. (1965). Partial reinforcement and breadth of learning. *J. exp. Psychol.* (in press).

TINBERGEN, N. (1952). "Derived" activities: Their causation, biological significance, origin, and emancipation during evolution. *Quart. Rev. Biol.*, **27**, 1–32.

WEINSTOCK, S. (1958). Acquisition and extinction of a partially reinforced running response at a 24 hr. intertrial interval. *J. exp. Psychol.*, **56**, 151–8.

ZEIGLER, H. P. (1964). Displacement activity and motivational theory. A case study in the history of ethology. *Psychol. Bull.*, **61**, 362–76.

Manuscript received 20th May, 1965.

CONTINUED WORD ASSOCIATIONS AND FREE RECALL

BY

W. A. MATTHEWS*

From the Department of Psychology, The University, Hull

Two experiments on the short-term free recall of 12-word associated and non-associated lists are reported. Degree of association (derived from norms obtained by continuous controlled association) and word frequency were varied. Significant facilitation as a result of the associative manipulations was obtained and clustering of the responses was positively related to this. Clustering was also affected by the method of presentation of the associated words; this occurred more often when they were grouped in presentation than when they were presented randomly arranged among other words in the list. Low frequency associated word lists were generally found to be more efficiently recalled than those of comparable association values but consisting of high frequency words.

Introduction

Various experiments have related the efficiency of free recall to the existence of categorical relationships or to associative connections between the items making up the lists (e.g. Bousfield, 1953; Jenkins and Russell, 1952; Deese, 1959). The investigators who had association measures as their independent variable used the norms obtained, by Russell and Jenkins (1954), from a large population of subjects each of whom gave a single response to each stimulus. It may be argued that norms obtained in this way distort the relationship which these measures are assumed to have to some intra-subject verbal organization in that high commonality responses necessarily reduce the number of subjects available to produce other words as associates. Thus in these situations, an association with a low score in the norms may be weakly associated to its stimulus or may be strongly associated but with this incorrectly assessed by the method used. The failure of Jenkins (1964) to obtain differential performance in paired-associate learning as the degree of association differed may be ascribed to this. This objection has less force when responses are obtained by continued controlled association in which each subject produces as many associations as he wishes to the stimulus word over a period of time (in this case 2 min.). Norms were obtained to 28 stimuli in this way. These were not identical with the stimuli used by Russell and Jenkins, as no direct comparison was intended with the results of experiments using their norms; rather interest was centred on the performance in free recall. The normative score was obtained by taking the number of subjects (from the groups of 50 which produced the norms) who gave a particular response anywhere in the associations produced over a 2 min. period. Thus, for example, the word "left" was given 32 times (i.e. by 32/50 subjects) to "right," whereas "arm" was only given four times. Using these norms, Matthews, Marcer and Morgan (1964) constructed 12-word lists, which consisted of four sets of three associated words. Of these three-word groups, one word was one of the 28 stimuli (a "core" word) and the other two words were associations to it. The mean associative value of the list was obtained by taking the mean of the associative connections from the eight associations in a list to their respective core words. Using a high and a low level of associative frequency, significant facilitation over the free recall of random control lists was obtained, and correlated with this was an increasing incidence of intra-list organization into associated groups. The

* Now at Department of Psychology, University of Southampton.

following experiments extend these earlier investigations by sampling four points on the associative-frequency dimension and introduce a word-frequency variable (Thorndike and Lorge, 1944) into the design. Additionally, in an attempt to isolate other important factors in the situation, further associative norms were obtained for 62 other words.

Standardization Procedure

Stage 1. Twenty-eight words were selected and presented to subjects both visually and auditorily. Each word consisted of five letters; and two different words within the group began with the same letter of the alphabet. The subjects were requested to associate for 2 min. to each of these words. The associations were written on mimeographed response sheets. With one exception ("docks") each stimulus occurred more than 50 times per million words sampled (Thorndike and Lorge G-count, 1944). It must be emphasized that controlled (and not chained) association was required, i.e. associations directly derived from the stimulus; and to facilitate this, the stimulus was written on the response sheet before association began. Examples with different stimuli preceded each session. Response repetition was allowed but seldom occurred.

Stage 2. A further 62 words were presented in a similar way for a 2-min. association period, with similar instructions as in stage 1 except that, at the end of 1 min., the association which had just been produced was underlined. These words consisted of 32 words occurring less than 50 times per million words sampled (in future labelled LF words) and 30 words occurring more often than 50 times per million words sampled (in future labelled HF words). They were associations given to "core" words in a sample of the 12-word lists used. These were given in four groups on different occasions with an approximately equal number of HF and LF stimuli presented randomly in each session.

Subjects. The subjects were five groups of 50 training college students (average age 20 years). Each group contained 25 men and 25 women. Four of the groups took part in stage 2 of the standardization procedure. The other group took part in stage 1. The subjects were not grouped for reasons which may have produced differential word association performance. The associations were summarized irrespective of order of emission, and normative data for each stimulus was produced. The norms consisted of the total number of instances each associate was given to each stimulus. These ranged from 1 to an actual maximum of about 50. Separate classifications were made for men and women but were combined for use in these experiments.

Experimental investigations

Three experiments were carried out but only two will be reported in detail here. In Experiment 1, the 12-word lists were presented with the four sets of three associated words in sequence. In the two other experiments, the same lists were used but the associated triplets were broken up and randomly presented among the other words in the list. In the first of these, training college students were used, as in the norm producing group; in the second, the subjects were university undergraduates. In the latter experiment, six presentations of each list were made whereas only three presentations were possible with the training college students. The greater amount of data produced by the university groups confirmed and extended the results of the comparable training college experiment, but the differences in the population from which the subjects were drawn make the randomized presentation to training college students more directly comparable with the ordered presentation of Experiment 1. This will be reported in detail but confirming results from the excluded experiment will be mentioned where appropriate.

Experiment 1

Material. Forty-eight 12-word lists were constructed from the data of stage 1 of the standardization test and the Thorndike-Lorge G-count. These were in four sets of 12. Each set of 12 consisted of four conditions; three HF associated lists, three LF associated lists, three HF control lists and three LF control lists. Each associated list consisted of four core words (i.e. words used as stimuli in stage 1 of the standardization test) and two associates to each core word. The control lists were of words which appeared to be not associated. They were matched for word frequency with an appropriate associated list. Initial letter duplication within a list was avoided and word length kept as constant

as the available material allowed. The four sets of lists were distinguished by having different mean associative strengths. The four mean values were approximately 22, 14, 7 and 2 (i.e. each associate was given to its core word by an average of 22/50 subjects, 14/50 subjects, etc., in the standardization group). A sample of these lists appears in Table I along with the associative norm score and G-count frequency.

TABLE I

A SAMPLE OF THE ASSOCIATED LISTS AT EACH ASSOCIATIVE AND WORD FREQUENCY POINT USED IN EXPERIMENTS 1 AND 2

AF22 HF			*AF14 HF*			*AF7 HF*			*AF2 HF*		
	AF	*WF*		*AF*	*WF*		*AF*	*WF*		*AF*	*WF*
right	C	AA	thing	C	AA	house	C	AA	march	C	AA
hand	23	AA	any	14	AA	dog	7	AA	order	3	AA
left	32	AA	some	12	AA	people	6	AA	journey	2	A
front	C	AA	water	C	AA	right	C	AA	cabin	C	A
back	34	AA	bath	17	46	time	5	AA	forest	8	AA
door	32	AA	reservoir	12	8	yes	5	AA	sleep	4	AA
march	C	AA	needs	C	AA	laugh	C	AA	right	C	AA
army	27	AA	clothes	18	AA	cry	7	AA	arm	4	AA
wind	27	AA	desire	11	AA	mouth	6	AA	path	2	AA
thing	C	AA	light	C	AA	field	C	AA	house	C	AA
object	26	AA	electric	16	AA	work	6	AA	boat	4	AA
some	12	AA	fire	12	AA	soil	5	AA	live	4	AA
AF22 LF			*AF14 LF*			*AF7 LF*			*AF2 LF*		
	AF	*WF*		*AF*	*WF*		*AF*	*WF*		*AF*	*WF*
field	C	AA	docks	C	16	cloth	C	A	point	C	AA
buttercup	18	2	cargo	20	9	fabric	6	21	cliffs	2	38
plough	15	17	warehouse	13	10	weave	5	22	mathematics	2	8
laugh	C	AA	field	C	AA	music	C	AA	laugh	C	AA
joke	30	13	buttercup	18	2	overture	3	4	tickle	3	10
giggle	9	5	plough	15	17	trumpet	6	17	sarcasm	4	3
music	C	AA	guard	C	AA	night	C	AA	giant	C	A
orchestra	20	10	rifle	7	31	pillow	5	33	ugly	4	34
violin	19	11	sentry	12	5	bats	4	19	footprints	2	3
docks	C	16	night	C	AA	guard	C	AA	doubt	C	AA
cranes	31	19	owl	15	35	van	6	29	anxiety	2	22
warehouse	13	10	light	10	AA	sentry	12	5	query	5	9

Procedure. Each list was presented three times in the learning session but these presentations were not sequential. Within each associative set, the 12 different lists were presented in a randomized order. The 12 lists were then re-randomized, and presented. This occurred a second time for the final presentation. The number of interpolated lists between repeated presentations was kept as similar as possible with the constraint that a pattern must not develop. Where the number of interpolated lists varied to meet the last constraint the overriding criterion that no list should appear twice in succession was always met. Within each list each triplet varied in position from trial to trial, occurring in either the first, second, third or last quarter. Due to practical difficulties, four presentations of each list were not possible: but as each list, in each condition, in each associative set of the experiment, varied in the same way, comparisons between lists and between associative sets were possible. Within the triplets, the order of the items was varied on each presentation with the core words preceding the associations once only.

This complicated and rather unusual randomization procedure was used to provide information on two aspects of the subjects' behaviour, i.e. the number of items recalled, and the way these items were clustered. Despite the instructions that free recall was allowed it was thought that normal sequential presentation would contaminate the results and serial learning be shown. Additionally, as the number of "core" words and associations from stage 1 was limited, an extension of list length was not possible, and so a "ceiling" effect was likely which might obscure differences between the four associative conditions (see Matthews, Marcer and Morgan, 1964).

The complete series of lists for each associative set were recorded on magnetic tape and presented auditorily at a constant speed of 1 word/sec. Each list was preceded by a warning signal (a low pitched buzz) which lasted for 1 sec. The lists were read in a regular monotone (male voice) beginning 1 sec. after the termination of the warning signal. Immediately after the twelfth word had been read, recall was made in writing

on prepared response sheets. Recall was free. The recall interval was approximately 45 sec. As performance improved, slightly more time was required but this never exceeded 60 sec. Reference back to earlier lists was not allowed. Before each series began, practice lists were presented during which some knowledge of the procedure was acquired and signal levels were adjusted so that all subjects could hear clearly.

Experiment 2

Material. This was the same as in Experiment 1.

Procedure. This was identical with Experiment 1 except for the following variation in within-list organization. The associated triplets were split up in this experiment and the 12 items were arranged randomly with the restriction that two sequential items should not be associated. The three-word sequences in positions 1–3, 4–6, 7–9 and 10–12 of the list were then treated as in Experiment 1 with the within-list positions and orders changing on each presentation trial. Although ceiling effects were thought to be unlikely with this intra-list arrangement, the pattern of list presentation as in Experiment 1 was maintained to make comparisons between the results more reasonable.

Subjects. These were men and women drawn from the teachers' training college at which the associative norms had been obtained, but these had not been involved in the norm production sessions. They were naïve to verbal learning experiments. In Experiment 1, there were 10 (12) subjects in the AF22 group, 12 (12) subjects in the AF14 group, 12 (10) subjects in the AF7 group, and 10 (7) subjects in the AF2 group. The figures in brackets give the number of subjects taking part in Experiment 2. These groups were selected at random from the available student population.

Results

The mean number of items correct on presentations 1 and 3 within each associative set is shown in Table II. Responses which were clearly errors of hearing or were homophones were treated as correct. If there was any doubt they were treated as errors.

TABLE II

Number of Items Correct on Presentations 1 and 3 in Experiments 1 and 2

		Associative HF condition		*Associative LF condition*		*Control HF condition*		*Control LF condition*	
		Expt. 1	*Expt. 2*	*Expt. 1*	*Expt. 2*	*Expt. 1*	*Expt. 2*	*Expt. 1*	*Expt. 2*
AF22	P1	22·6	17·4	21·6	14·8	17·2	16·5	13·9	14·8
	P3	26·1	19·6	27·9	22·0	18·7	19·3	18·5	18·2
AF14	P1	23·3	17·5	26·3	16·0	17·9	14·8	18·0	15·8
	P3	29·0	20·8	31·6	21·0	19·3	16·1	20·3	17·0
AF7	P1	23·8	19·0	25·5	18·8	17·6	15·5	15·0	15·7
	P3	24·9	20·9	26·1	22·5	18·1	18·9	15·9	20·5
AF2	P1	22·3	15·4	20·9	15·4	16·9	14·1	16·5	11·8
	P3	23·8	19·8	25·3	19·4	20·4	17·0	20·3	16·4

Bartlett tests were carried out, and the single transformation required was made. Analyses of variance (repeated measures model) were then carried out on the number of correct responses per presentation for each associative set separately. A selection of these results will be discussed. In Experiments 1 and 2, all the differences between the associated and the control lists were significant ($p < 0·001$) with the single exception of AF22 lists in Experiment 2. Also large inter-subject differences occurred in the general levels of performance in this situation (p values ranged from $<0·05$ to $<0·001$). To elucidate the degree to which recall was facilitated as a function

of the level of associative frequency in Experiment 1, the control scores of nine subjects from each associative set were matched and an analysis of variance carried out on this set of scores. This repeated the findings from the separate analyses of variance above; and the residual variance was used to carry out a Duncan multiple range test (Edwards, 1960) to evaluate whether any of the differences between the 16 mean scores (eight associated with eight control conditions) were significant at the 0·05 level. The control lists did not differ among themselves, although four of the six highest means were from HF control lists, but were all significantly less well recalled than the associated lists. Among the latter lists an interesting relationship was found. The three AF conditions (22, 14 and 7) with low frequency (LF) words as the associates produced more correct responses than in the comparable high frequency (HF) lists. The AF14 LF lists were significantly better recalled than any other lists. No significant differences were found between the AF22 LF and AF7 LF lists and the AF14 HF and AF22 HF lists. The former two lists were significantly different from the AF7 HF lists. Both the AF2 lists were inferior to the other associated lists, but only the AF2 HF lists were significantly worse than the AF7 HF lists. This superiority of the LF lists occurred consistently in Experiment 1 and was beginning to develop by the third presentation in Experiment 2 (except in the AF2 set). It was clear by trial 6 in the experiment not reported in detail here.

The differences which appeared clearly in Experiment 1 may be traced to the incidence of clustering into three-word groups. This did not appear in Experiment 2. These results are shown in Table III.

TABLE III

MEAN NUMBER OF TRIPLETS OCCURRING IN ASSOCIATED AND CONTROL LISTS IN EXPERIMENTS 1 AND 2

	Experiment 1				*Experiment 2*			
	Associated		*Controls*		*Associated*		*Controls*	
	HF	*LF*	*HF*	*LF*	*HF*	*LF*	*HF*	*LF*
AF22	14·0	17·0	5·3	4·7	0·6	1·6	0·2	0·2
AF14	18·5	21·3	6·4	5·8	1·8	1·3	0·3	0·3
AF7	11·2	15·3	6·8	4·6	0·7	2·1	0·1	0·0
AF2	11·5	13·7	5·9	6·1	0·6	1·1	0·1	0·1

In Experiment 1, it is clear that a greater proportion of the LF associated lists than of the HF associated lists occur as three-item sequences, and this appears to be related to the number of words correctly reproduced. To clarify this, the number of triplets produced by each subject was ascertained. These were adjusted to allow for an increasing probability of occurrence of triplets by chance as the number of items correctly reported, increased (for details of method, see Matthews, Marcer and Morgan, 1964). A rank order correlation (Spearman's rho) was then carried out between these adjusted scores and the number of items correctly reported by each subject under each condition in each associative set. A summary of these appears in Table IV. The small number of triplets in Experiment 2 prevented this analysis being carried out. The consistently higher correlation for the associated LF than the associated HF lists (with identical subjects at the four different associative levels) indicates that the relationship is not artefactual.

TABLE IV

RANK ORDER CORRELATIONS (SPEARMAN'S RHO) BETWEEN THE ADJUSTED NUMBER OF TRIPLETS REPORTED AND THE TOTAL NUMBER OF ITEMS CORRECT BY ASSOCIATIVE SET AND CONDITION IN EXPERIMENT 1

	HF associative	*LF associative*	*HF controls*	*LF controls*
AF22	0·79 (0·01)	0·92 (0·01)	−0·76 (0·05)	0·52 (NS)
AF14	0·64 (0·05)	0·91 (0·01)	0·56 (0·05)	0·48 (NS)
AF7	0·76 (0·01)	0·82 (0·01)	0·31 (NS)	0·38 (NS)
AF2	0·67 (0·05)	0·80 (0·01)	0·60 (0·05)	0·41 (NS)

The significance levels for the appropriate Ns appear in brackets.

To elucidate this finding, the associative responses to the 62 words in part 2 of the standardization tests were used. These stimuli were the associative responses used in eight of the associated lists (one HF and one LF list from each associative frequency level). From this data, an estimate was made of the extent to which the items in each list elicit each other as associates, and of the mean associative strength of these intra-list connections. The number of words which appeared as associates to one or more of the other words varied little as a function of either word or associative frequency. The mean associative strengths of these connections did vary, however, as a function of both associative level (which was partially inevitable) and word frequency; and the mean associative strengths by condition follow. In the AF22 HF list, the mean inter-item association value is 16·5 (i.e. the intra-list associations are produced by an average of 16·5/50 subjects); for the AF22 LF list the value is 25·9; AF14 HF is 9·9, AF14 LF is 17·00; AF7 HF is 9·89; AF7 LF is 13·09; AF2 HF is 3·7; and AF2 LF is 9·5. Thus the probability with which the intra-list associates elicited one another was always higher in the LF than in the HF lists. A final inference which may be drawn from Tables III and IV is that these LF associative interconnections were more likely to be within a triplet than are those in HF lists. On analysis, 7 from 83 (8·4 per cent.) LF words were given as associates to words outside the associated triplet, whereas 15 from 78 (19·2 per cent.) HF words were classifiable in this way.

Despite the varied methods of presentation, an increasing number of correct items was reliably found over the three presentation trials in both Experiments 1 and 2 ($p < 0·001$) with a single exception. However, the rate of improvement varied little between the experimental and the control conditions in Experiment 1 where none of the four association conditions × order of presentation interactions reached the 0·05 significance levels. In Experiment 2, two of these interactions were significant (AF22, $p < 0·05$; AF14, $p < 0·001$), indicating that the appearance of strongly associated words did produce increasingly efficient recall with a randomized type of presentation.

DISCUSSION

Associative facilitation was clear in both experiments, and this was greater in Experiment 1 where the associated words occurred infrequently in the language. However, an increasing level of associative connection did not lead to increasingly efficient performance in free recall. AF22 and AF14 lists were not differentiated, and AF7 lists were not clearly less well recalled than these two other sets. One inference from this is that the degree of overlearning of associative connections is

so great that differences in recall may only be found between high/medium and very low levels of associative connections in a free recall situation. The less efficient performance on the AF2 lists, where the associations had been given by only two from 50 subjects anywhere in their response protocols, supports this view.

Facilitation was clearly related to an increased incidence of clustering into triplets, and LF words were more clearly grouped than HF words in Experiment 1. This was appearing by the third presentation in Experiment 2 (and had appeared in the number of pairs reported). It was also found in the third experiment mentioned earlier. Postman (1964) has reported that HF words are more affected by interfering associations than are LF words and uses the explanatory notion of associative interference (Underwood and Schulz, 1960). It seems feasible that the number of associative groups to which HF responses may be connected will be greater than those to which LF responses are related. As recall is related to associative grouping, and if some uncertainty exists as to the correct response, then the greater number of alternatives which are possible to the HF words may reduce the efficiency of clustering and introduce incorrect, but associatively related, words into the responses. This should be less likely in LF lists. The stronger inter-item associative connections between the LF responses reported earlier supports this view and indicates that, despite equating on one-way associative norms, there were much stronger associative connections between the LF than the HF words. This is supplemented by the smaller percentage of LF than HF words which have associative connections outside the particular associative triplet (8·4 per cent. and 19·2 per cent., respectively). Thus the looser organization of HF associated lists may make the effective size of an associated group of words in an HF list larger than three words. This may both confuse some classifying system used to "label" each associative group and, less probably, exceed the number of items in a group which a subject can handle efficiently.

It seems possible to reduce the LF associative superiority to a process dependent on the frequency of experience of words occurring normally in few, rather than many, varied contexts; and not solely to similarities in the actual frequency of associates to a particular stimulus. This interpretation would be supported by an experiment in which the complete inter-item associations between the words in both HF and LF lists had been equated and no difference in recall or clustering found. However, such lists would be very difficult to produce. The importance of the clustering strategies which this interpretation assumes is further emphasized by the number of three-word clusters (as a function of presentation order) which appeared in the control lists. Where the structure of the list does not correlate with the associative groupings, clustering develops more slowly, cluster size is reduced, and recall is less efficient. In these conditions, the order and speed of presentation may be the major determinants of recall, and "chunking" (Miller, 1956) will not be so readily facilitated by the associative cues present.

Thanks are due to Dr. H. D. Wing, Principal, City of Sheffield Training College for helping to provide subjects.

References

Bousfield, W. A. (1953). The occurrence of clustering in the recall of randomly arranged associates. *J. gen. Psychol.*, **49**, 229–40.

Deese, J. (1959). Influence of inter-item associative strength upon immediate free recall. *Psychol. Rep.*, **5**, 305–12.

Edwards, A. L. (1960). *Experimental Design in Psychological Research.* New York: Holt, Rinehart and Winston.

Jenkins, J. J. (1964). The 1952 word association norms. In Postman, L. (Ed.), *Norms of Word Association.* New York: Academic Press.

JENKINS, J. J., and RUSSELL, W. A. (1952). Associative clustering during recall. *J. abnorm. soc. Psychol.*, **47**, 818–21.

MATTHEWS, W. A., MARCER, D., and MORGAN, E. (1964). Word association hierarchies and free recall. *J. verb Learn. verb Behav.*, **3**, 371–5.

MILLER, G. (1956). The magical number 7 plus or minus 2. *Psychol. Rev.*, **63**, 81–97.

POSTMAN, L. (1964). Acquisition and retention of consistent associative responses. *J. exp. Psychol.*, **64**, 183–90.

RUSSELL, W. A., and JENKINS, J. J. (1954). The complete Minnesota norms for responses to 100 words from the Kent-Rosanoff association test. *Tech. Rep. No.* 11 *Univ. of Minn. ONR*, Contract N8 ONR 66216.

THORNDIKE, E. L., and LORGE, I. (1944). *The Teachers' Word Book of* 30,000 *Words.* New York: Bureau of Publications, Teachers College, Columbia University.

UNDERWOOD, B. J., and SCHULZ, R. W. (1960). *Meaningfulness and Verbal Learning.* Philadelphia: Lippincott.

Manuscript received 7th May, 1965.

VIGILANCE–AROUSAL vs. REINFORCEMENT

BY

HANS CHRISTOPH MICKO

From Göttingen University

Task-irrelevant stimuli (projected jokes, which were difficult to read) received an increasing attention as the auditory vigilance session progressed. This result supports reinforcement theories at the cost of activation theory of vigilance.

Introduction

Several theoretical approaches have been developed in order to explain and predict performance in vigilance tasks. Some of these could be classified as *Reinforcement theories*, e.g. Broadbent's (1958) filter theory, Claridge's (1960) and Treadwell's (1960) discussion of vigilance performance in Hullian terminology, Holland's (1958) Skinnerian approach and Mackworth's (1950) slightly different interpretation of vigilance deterioration as secondary extinction or generalized inhibition of the responding behaviour. According to these theoretical interpretations, which are basically alike, vigilance level, or detection probability, is a function of the frequency and recency of reinforcement of the watch-keeping behaviour by signal detections. Vigilance deterioration is assumed to be due to the rareness of reinforcements, since signals which can be detected are infrequent in vigilance experiments (Schmidtke and Micko, 1964).

A different interpretation is presented by the *Activation theory*, which was first mentioned by Deese (1955) and discussed later among others by Broadbent (1958, 1963), and Haider (1962). According to the activation theory, vigilance level is a function of reticular activity. Reticular activity itself depends on the variability of the environmental stimulation, which is typically low in vigilance experiments. This prolonged low stimulus variability of the experimental situation is assumed to be responsible for the deterioration of watch-keeping performance.

Of these two theories the activation theory is the more general one. From it one is able to predict all data supporting reinforcement theories, by interpreting detected signals as variations of the stimulus environment. (We shall not be concerned here with other interpretations, e.g. expectancy theory, etc.)

Most vigilance studies have been interpreted as supporting one or other theory. McGrath (1963) devised the only crucial experiment known to the author. He argued that, according to activation theory, distracting stimulation ought to improve detection performance, whereas according to reinforcement theories distractions should have an adverse or no effect. In fact, McGrath found that distraction interacts with the difficulty of the vigilance task in influencing detection performance.

Aims

The present critical experiment was provoked by a reasoning different from that of McGrath's but it leads to an experimental arrangement similar to his. If deterioration of vigilance is due to a decrease of arousal, we must assume that all organized mental activity decreases during the watch-keeping session. From the "Lack-of-reinforcement" theory we should predict that only the watch-keeping behaviour undergoes extinction. Other, task-irrelevant, mental activities may to

some extent undergo extinction by the mechanism of generalized inhibition but to a much lesser degree than the watch-keeping behaviour. Therefore task-irrelevant activites, e.g. thinking about the last date or the coming examination will become more and more frequently dominant and will for a while replace the observing behaviour.

We cannot judge from a subject's detection performance, if periods of non-attendance are filled by drowsiness and light sleep or by distracting thoughts or fantasies. Introspective reports are conflicting and inconclusive (McGrath, 1960; Bakan, 1963). Therefore, in order to make mental activities other than the required watch-keeping observable, one may supply the subjects with task-irrelevant stimuli, which are to compete with the task stimuli in a similar manner to other, mainly internal, stimuli such as memories, ideas, etc.

Such extra stimuli have to fulfill three requirements: (i) They must be the most likely ones to replace the observing behaviour, if it is replaced at all. (ii) They must not distract the subjects from the task stimuli more than do other internal or external stimuli. (iii) Attending to these stimuli should require a certain amount of effort and "organized mental activity," since the arousal theory would admit the replacement of watch-keeping by disorganized mental activities, such as are reported for states intermediate between waking and sleeping (Oswald, 1962).

Task-irrelevant stimuli, which meet these requirements should, according to the reinforcement theories, receive an increasing attention in the course of the watch-keeping session, whereas according to the activation theory, they should be attended to in a decreasing fashion.

Method

Vigilance task. A series of 1000 c-p.s. tones lasting 0·2 sec. was recorded on tape together with white noise of medium intensity for masking purposes. Intervals between tones were 0·6 sec. Interspersed in this series were 24 tones of somewhat higher intensity, constituting the signals to be detected. The difference in intensity between critical and non-critical tones had been adjusted in pretests, so that 90 per cent. detections could be expected under alert conditions. The signals were so spaced that each of the following intervals occurred once every $\frac{1}{4}$ hr.: $\frac{1}{2}$ min., 1 min., 2 min., 3 min., 4 min., $4\frac{1}{2}$ min. Based on this principle critical signals were presented 3, 7, $7\frac{1}{2}$, $9\frac{1}{2}$, 14, 15, 17, 18, $22\frac{1}{2}$, $25\frac{1}{2}$, 26, 30, 32, $36\frac{1}{2}$, $37\frac{1}{2}$, $41\frac{1}{2}$, $44\frac{1}{2}$, 45, 48, $48\frac{1}{2}$, $52\frac{1}{2}$, $53\frac{1}{2}$, 58, 60 min. after the beginning of the session. Subjects were required to respond to the critical signals by pressing a button and thus switching on a lamp on the experimenter's desk in a separate room. This lamp remained burning until the experimenter switched it off.

Extra stimulation. Forty-five jokes of between 25 and 72 words length, randomly spaced, were projected on a screen, one every 80 sec. The diaphragm of the projector was turned out of focus, so that reading the print was possible but required considerable effort.

Subjects. Twenty-eight male and female students of Göttingen University volunteered as subjects. All reported normal eyesight (with or without spectacles) and normal hearing capacity. Of these 19 served as "listeners"—these are the ones of interest to us—and 9 as "readers."

Procedure. The task was presented as an experiment on distractability. Contrary to fact, "readers" considered themselves as members of the experimental group, while the "listeners" thought of themselves as controls.

The "listeners" were given the usual vigilance test instruction. They were not forbidden to read the jokes, since the experimenter declared that he was "unable and unwilling to control whether they read the jokes or not." However, subjects were told that it was their task to listen to the signals, and failure to pay full attention to the vigilance task would most probably result in the missing of signals.

"Readers" were employed (i) to have a control that the jokes were still readable, (ii) to give the "listeners" a plausible and harmless explanation for the projection of jokes.

Their task was the deciphering of the jokes. They were further instructed not to attend to the tones, but to count any critical signals of the vigilance task which they might hear by accident.

The experiment took place in a darkened lecture room. Ten cubicles (two rows of five) were constructed using cardboard walls, thus separating subjects completely from each other, but allowing them a good view of the screen. Each cubicle contained a fairly comfortable chair and a response button. Subjects were tested either in the mornings, afternoons or evenings of three successive days. In fact, only between three and seven subjects were tested simultaneously since it proved difficult to recruit more for any one particular time. At every session about one-third of those present served as readers and two-thirds as listeners. Watches were removed and subjects were asked not to smoke.

The experiment consisted of the following stages:

(i) Instruction and two 3-min. trial runs of the vigilance task, each containing four signals.

(ii) A 15-min. practice period in order to acquaint subjects with the low signal frequency and to adapt them to the illumination. The screen was illuminated during this period but no jokes were presented. All subjects were required to carry out the listening task.

(iii) A pause of approximately 5 min. duration, in which subjects were given an opportunity to stretch, stand up, and chat.

(iv) The 1-hr. test session.

(v) An unexpected "memory-test" in which the experimenter read the jokes which had been presented on the screen and some others, up to the punch line. All subjects had then to complete the jokes.

Results

Subjects reported that none of the jokes was known to them before the experiment; and that very few of those they attended to during the experiment were not remembered.

Biserial rank correlations (Whitfield, 1947) have been computed between the positions of the jokes in the joke-sequence and the dichotomy "remembered/not remembered" for each subject separately. Eight of the 19 correlations were significant on the 0·05 level. The trend curves representing these eight correlations are shown in Figure 1 (*a*). Seven of the significant correlations were positive, (the smallest) one negative. The prevailing of the positive over the negative correlations is significant (binominial test, $p < 0{\cdot}05$). With one or two exceptions the insignificant correlations came from subjects who read either most of or hardly any of the jokes. Reading trend curves for these subjects are given in Figure 1 (*b*). They were tested for non-randomness of the read/not read-sequence by the one sample runs test. None of them showed a significant deviation from randomness, the largest value obtained (for the U-shaped curve in Figure 1 (*b*)) being $z = 1{\cdot}26$, i.e. $p = 0{\cdot}21$.

Of the nine readers, one stopped reading after half an hour because of eye trouble, a second one did not read the last five jokes. The others read and remembered practically all jokes (one or another missed one or two jokes at various parts of the session).

A comparison of the vigilance performance was made between the following three groups:

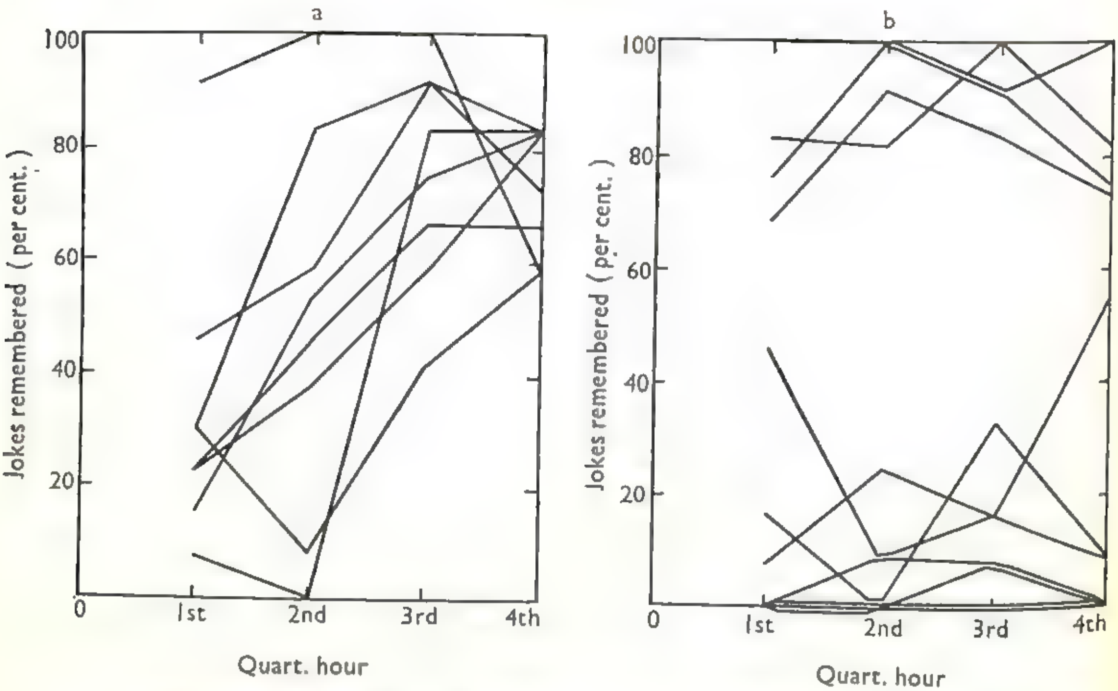
(i) The four subjects who read many jokes throughout.

(ii) The seven subjects who read few jokes throughout.

(iii) The seven subjects who read few jokes at the beginning and many at the end.

The performance curves are given in Figure 2. No significant difference could be ascertained for average performance (Kruskal-Wallis-H $= 0{\cdot}86$, $p > 0{\cdot}50$) or

FIGURE 1



Attendance to task-irrelevant stimuli as a function of time (main experiment). Each line represents one subject.

(*a*) Significant trends.

(*b*) Insignificant trends.

TABLE I

PROPORTION OF SUBJECTS DETECTING EACH CRITICAL STIMULUS

	Number of signal in the sequence											
	1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*
A	1·00	1·00	1·00	1·00	1·00	1·00	0·80	0·43	0·63	0·86	1·00	0·80
B	0·74	1·00	1·00	0·85	0·86	0·93	0·93	0·50	0·73	0·58	0·91	0·89
C	0·26	0	0	0·15	0·14	0·07	−0·13	−0·07	−0·10	0·28	0·09	−0·09

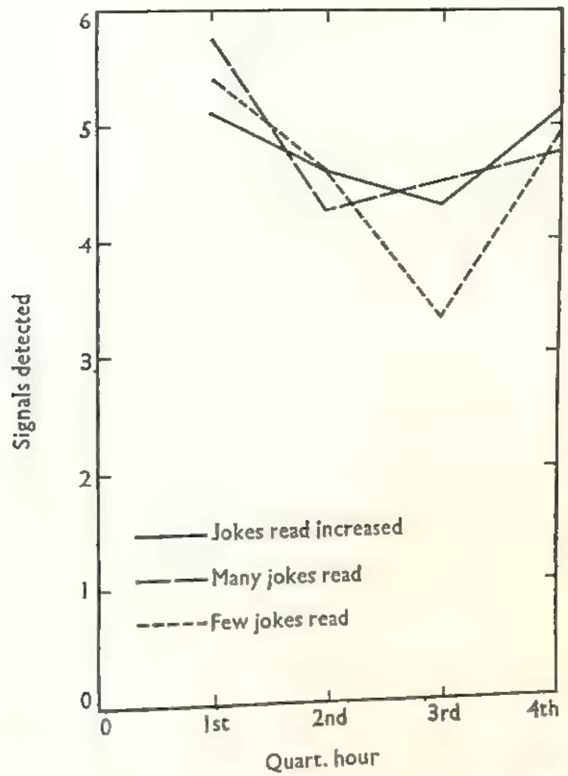
	Number of signal in the sequence											
	13	*14*	*15*	*16*	*17*	*18*	*19*	*20*	*21*	*22*	*23*	*24*
A	0·89	0·64	0·75	0·63	0·60	1·00	0·78	0·90	1·00	0·69	0·86	0·86
B	0·90	0·40	0·43	0·25	0·45	0·80	0·90	0·78	0·86	0·50	0·75	0·88
C	−0·01	0·24	0·32	0·38	0·15	0·20	−0·12	0·12	0·14	0·19	0·11	−0·02

A: Proportion of subjects reading the simultaneously presented joke, who detected the signal.

B: Proportion of subjects not reading the simultaneously presented joke, who detected the signal.

C: Difference between A and B.

FIGURE 2



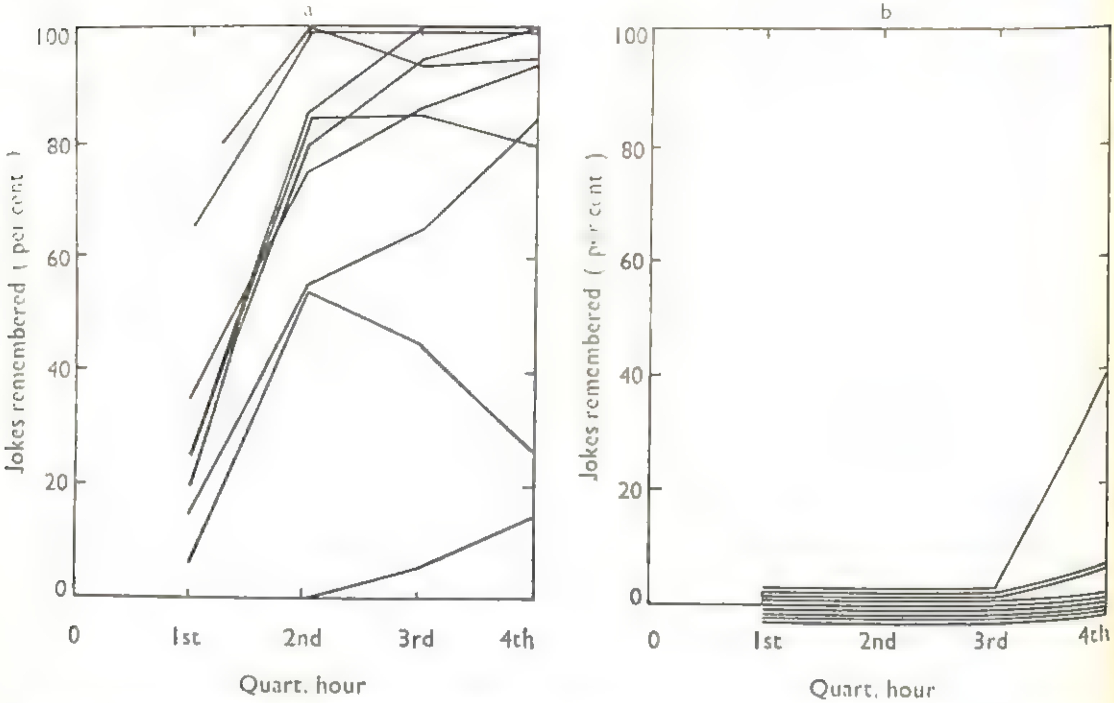
Average vigilance performance of subjects who read more jokes at the end than at the beginning (solid line), subjects who read many jokes throughout (long dashes) and those who read few jokes throughout (short dashes).

for performance decrement (the "most nearly significant" result obtained being for the difference between "performance in the first half-hour and performance in the second half-hour: Kruskal-Wallis-H = 2·0, $p > 0·30$). The overall decrement from the first to the second half-hour was significant (Wilcoxon-T = 16·5, $p < 0·01$). Table I gives for each critical signal the detection rate of subjects who read the joke presented at the same time and of those who did not read it. Subjects who read the joke just presented gave a superior performance on 15 of the signals and an inferior performance on seven of them. However, none of the differences approached significance. An overall test is impossible, because the measurements are partly dependent. The sequence of differences given in the last row of Table I did not show a significant trend except a short-wave fluctuation which must remain unexplained although it seems to be non-random (if the second difference is considered slightly larger than the third, the binominial trend test (Moore and Wallis, 1943) gives a $p = 0·019$, if it is considered slightly smaller a $p = 0·086$ is obtained).

Additional evidence

A similar experiment was conducted several months before on 18 "listeners" and 12 "readers." Vigilance test, jokes, and procedure were the same, but there was no 15-min. practice period and subjects were not separated by partitions. For half of the subjects, reading the jokes was fairly easy, for the other half very difficult indeed—so that 25–50 per cent. of the jokes were not deciphered even by the "readers."

FIGURE 3



Attendance to task-irrelevant stimuli as a function of time (preliminary experiment). Each line represents one subject.

(*a*) Reading relatively easy.
(*b*) Reading very difficult.

Seven significant ($p < 0{\cdot}05$) biserial rank correlations between joke position and the "remembered/not remembered" dichotomy have been obtained. All of them were positive. This result is again significant (binominial test, $p < 0{\cdot}01$).

Six of the significant correlations stem from subjects who could read the jokes easily, i.e. all but the uppermost and the two lowermost trend curves of Figure 3 (*a*). One of the significant correlations stems from a subject to whom reading was difficult, i.e. the uppermost trend line of Figure 3 (*b*).

Average vigilance performance did not show a decrement. This may be due to the fact that subjects were not isolated well enough, so that extra cues could be obtained. Therefore not much weight should be given to vigilance performance data in this part of the experiment. Subjects to whom reading was difficult and those to whom reading was easy did not differ in vigilance performance. Neither was the probability of a subject detecting a signal affected by his reading the joke presented simultaneously with the signal.

Discussion

Since reading the jokes was neither suggested nor forbidden to the subjects (i.e. "listeners") and since the reading was interesting but also exerting it was to be expected that some subjects would decide to read all jokes and others none. From their behaviour we cannot learn which of our two competing predictions is correct. It was hoped, however, that one would find subjects whose attendance to task-irrelevant stimuli either increased or decreased during the session. These

would be the subjects of interest. In fact, all but one subject read more jokes towards the end of the session than in the beginning. This result supports the reinforcement theories at the cost of the activation theory. Of the reinforcement theories Broadbent's filter theory states most explicitly, that irrelevant information is selected at the cost of the task relevant information in an increasing rate in the course of the watch-keeping session.

As was expected, the easier it was to read the jokes, the earlier and the more frequently subjects shifted their attention to these task-irrelevant stimuli.

McGrath (1963) gave his subjects an album of photographs to look at while performing an auditory vigilance task. They reported a decrease of attending to the album from about two-thirds of the time in the first quarter of an hour to one-third of the time in the two intermediate periods and one-seventh in the last 15 min. This result is just the opposite of ours, and induced us to repeat the "preliminary experiment" (see additional evidence) providing for an adaptation period. Having obtained the same results as before, we do feel sure that the lack of adaptation was not responsible for our subjects not having started to read the jokes right from the beginning. Neither can we attribute such a great discrepancy of the result to McGrath's subjects being unable to judge the time spent looking at the photographs. Our best guess is, therefore, that McGrath's photographs attracted the subjects' attention more than our hardly readable jokes—in contrast to internal stimuli, which are the only source of distraction in typical vigilance experiments. This may have made subjects look at the photographs right from the beginning, especially since this was at least implicitly suggested by the instructions. McGrath's subjects apparently lost interest in the photographs fairly quickly, while ours, having been supplied with jokes only occasionally did not. Even looking at photographs, when performed in massed practice, may become inhibited and will then be supplemented by day-dreaming or other activities. This explanation of McGrath's result would be in line with the reinforcement theories, while it would be more difficult to explain the results of the present experiment, if one assumes a reduction of arousal level during the watch-keeping session.

An alternative interpretation would be, that in the present experiment the effect of the warning against the missing of signals, due to the distraction by the jokes, had dissipated as the watch progressed. This would, however, be the case with the implicit or explicit warnings against diverting attention towards internal stimulation given before every vigilance task.

In this experiment subjects who attended to the irrelevant stimuli, at any time during the session did at least as well in the vigilance task as those who did not attend to the irrelevant stimuli. Therefore we cannot say that attendance to distracting stimuli impairs detection performance, although detection performance deteriorated and attendance to distracting stimuli increased as the watch progressed. The following three explanations suggesting the need for further research could be presented to explain this fact.

(1) Subjects who did not attend to the jokes at all might nevertheless have occasionally attended to other irrelevant internal or external stimuli. Since the jokes have been attended to in an increasing fashion, the most plausible assumption would be that attendance to other irrelevant stimuli also increased with time.

(2) The jokes were projected for a longer time than subjects needed to read them. Those who read the jokes might have compensated their attendance deviations by increased efforts to detect signals in the intervals, i.e. reading the jokes may have resulted in either increased activation or dissipation of inhibition, increased novelty

of the task stimuli, etc., and consequently in increased vigilance performance. Presumably this interpretation would be favoured by Broadbent (McGrath, 1963, discussion).

(3) Subjects might be able to learn in the course of the vigilance session to attend to relevant as well as irrelevant stimuli at the same time, i.e. they might learn to detect critical signals while being partly occupied with other mental activities. This assumption, however, would necessitate a thorough rethinking of the theory of vigilance performance and is better discarded until supporting data have come forward.

The author wishes to thank Dr. R. Conrad for his valuable comments and constructive criticism.

References

Bakan, P. (1963). An analysis of retrospective reports following an auditory vigilance task. In Buckner, D. N., and McGrath, J. J. (Ed.), *Vigilance: a symposium.* New York: McGraw Hill.

Broadbent, D. E. (1958). *Perception and Communication.* London: Pergamon.

Broadbent, D. E. (1963). Possibilities and difficulties in the concept of arousal. In Buckner, D. N., and McGrath, J. J. (Ed.), *Vigilance: a symposium.* New York: McGraw Hill.

Claridge, G. (1960). The excitation-inhibition balance in neurotics. In Eysenck, H. J. (Ed.), *Experiments in Personality,* Vol. II. London: Routledge & Kegan Paul.

Deese, J. (1955). Some problems in the theory of vigilance. *Psychol. Rev.*, **62**, 359–68.

Haider, M. (1962). *Ermüdung, Beanspruchung und Leistung.* Vienna.

Holland, J. G. (1958). Human vigilance. *Science,* **128**, 61–7.

McGrath, J. J. (1960). Subjective reactions of vigilance performers. Supplementary note to *Tech. Rep. 2 Human Factors Research, Los Angeles* **19**.

McGrath, J. J. (1963). Irrelevant stimulation and vigilance performance. In Buckner, D. N., and McGrath, J. J. (Ed.), *Vigilance: a symposium.* New York: McGraw Hill.

Mackworth, N. H. (1950). Researches on the measurement of human performance. *Med. Res., Council spec. rep. Ser.*, No. 268.

Moore, G. H., and Wallis, A. W. (1943). Time series sign tests based on signs of differences. *J. amer. Statist. Ass.*, **38**, 514–00.

Oswald, I. (1962). *Sleeping and Waking.* Amsterdam: Elsevier.

Schmidtke, H., and Micko, H. C. (1964). *Untersuchungen über die Reaktionszeit bei Dauerbeobachtung.* Cologne.

Treadwell, E. (1960). The effects of depressant drugs on vigilance and psychomotor performance. In Eysenck, H. J. (Ed.), *Experiments in Personality,* Vol. I. London: Routledge & Kegan Paul.

Whitfield, J. W. (1947). Rank correlation between two variables, one of which is ranked, the other dichotomous. *Biometrika,* **34**, 292.

Manuscript received 5th July, 1965.

NEGATIVE RECENCY AND EVENT-DEPENDENCE

BY

A. H. WINEFIELD

From the Department of Psychology, University of Adelaide

In a two-choice situation with event-probabilities of 0·67 and 0·33, recency effects were studied as a function of (*a*) event-dependence and (*b*) fore-knowledge of event-probabilities. The following conclusions are drawn: (1) Belief in some kind of event-dependence is a necessary condition for the occurrence of negative recency; (2) Fore-knowledge of the event-probabilities does not affect recency phenomena; (3) The behaviour of subjects in the two-choice situation, is less irrational than some writers have claimed.

Introduction

One of the most commonly observed sequence effects in the two-choice, probability-learning experiment concerns the reaction of subjects to homogeneous runs of one event. The typical reaction is to exhibit a decreasing tendency to predict the event as the run of it increases. This phenomenon was first reported by Jarvik (1951), who called it "negative recency" and compared it with the well known "gambler's fallacy." However, it seems unlikely that negative recency in general can be attributed to belief in the gambler's fallacy. As Cohen (1960) points out, ". . . even mathematicians who are perfectly convinced of the independence of the outcomes of successive tosses of a coin are still inclined to predict a particular outcome just because it has not occurred for a relatively long time in a series of tosses."

When subjects display negative recency they are behaving as if events are not occurring independently of one another but as if the occurrence of an event decreases the likelihood that it will occur on the subsequent trial. A model fulfilling this requirement would involve a finite population from which events are sampled without replacement. For example, if subjects were required to predict the outcome of successive drawings of balls from an urn containing red and blue balls when balls were not replaced after being drawn, it would be quite rational for them to display negative recency. In this case the drawing of a red ball on the *nth* trial does reduce the probability that a red ball will be drawn on the $n + 1th$ trial. (If there are N balls in the urn at the *nth* trial, R of which are red, then the probability of a red ball being drawn on the $n + 1th$ trial is reduced from R/N to $R\text{-}1/N\text{-}1$ by the drawing of a red ball on the *nth* trial.)

An important factor in this situation is whether or not subjects know the proportion of red to blue balls in the urn at the outset, and whether or not they know the total number of balls in the urn. Given this information subjects could always know the proportion of red to blue balls left in the urn and always choose the colour which predominated at any trial. However, if they did not know the proportion of red to blue at the outset, or if they did not know the total number of balls in the urn they could not adopt such a strategy. One would expect them to display negative recency only if they knew the proportion of red to blue balls at the outset, but did not know the number of balls in the urn. If they did not know the proportion of red to blue one would not necessarily expect negative recency. In that case they would almost certainly attempt to estimate the relative proportion of red to blue on the basis of what was drawn. Consequently, if a red ball were drawn on the *nth* trial this might increase their tendency to predict red on the $n + 1th$ trial since it would increase their estimate of the proportion of red balls in the urn. This would lead to

positive recency. However, the extent to which their estimate would be affected by a single drawing should decrease with the numbers of drawings. For example, suppose three balls only have been drawn, two red and one blue. One estimates that the proportion of red balls in the urn is two-thirds. Whatever is drawn on the fourth trial will change this estimate either to three-quarters (if a red ball is drawn) or to a half (if a blue ball is drawn). On the other hand, if one estimates that the proportion of red balls is two-thirds on the basis of several hundred drawings, one's estimate will scarcely be altered by a single drawing. Consequently, if subjects do not known the relative proportions of red and blue balls in the urn, one might expect them to display positive recency which decreases as the number of drawings increases. However, this is not the only reason for expecting positive recency to decrease. For in the same way as successive drawings affect one's estimate of the relative proportions less and less, they affect the actual relative proportions more and more because the number of balls in the urn is steadily decreasing. Thus the drawing of a red ball on a late trial will decrease the likelihood of a further red more than it will on an early trial, because there are fewer balls left in the urn. Thus the tendency for subjects to display negative recency should increase with successive drawings.

In short, if subjects know the relative proportions of red and blue balls in the urn one would expect them to display negative recency to an increasing extent; if they do not know the proportions of red and blue one would expect recency effects to change in the positive-negative direction with successive drawings. Of course, these predictions are based on purely rational considerations and there is very little evidence that subjects behave rationally in two-choice probability-learning situations. Nevertheless, they do provide some sort of criteria with which to evaluate behaviour.

The suggestion that negative recency is due to belief in the gambler's fallacy or to similar fallacious ideas about randomness and is therefore most likely to occur in a situation where events occur independently, is somewhat paradoxical in view of the fact that negative recency involves subjects behaving precisely as if events were *not* occurring independently. In fact, one is not convinced that negative recency has been demonstrated in a situation where subjects genuinely believe events to be occurring randomly. Subjects tend to be highly suspicious in most psychological experiments and frequently suspect that the experiment's covert aim is to test their intelligence. Consequently, it is much easier to convince them that they are required to solve a problem than it is to convince them that they are required to anticipate events in a random series. Hyman and Jenkins (1956) experienced this difficulty: "We found it easier to convince our subjects that the sequence of colours had a definite structure than to convince them that it was random."

Great care must be taken in order to convince subjects that events really are occurring randomly. Probably the best technique is to use a device which enables the subject to create his own random sequence. This technique was adopted by Morse and Runquist (1960) whose subjects dropped a rod on a pattern of parallel lines and predicted before each drop whether or not the rod would come to rest in contact with a line. Unfortunately, they were interested in probability-matching and do not report whether or not their subjects displayed negative recency.

In the following experiment recency effects are studied as a function of event-dependence and of fore-knowledge of the event-probabilities. The hypothesis was that negative recency only occurs when subjects believe events to be dependent and that knowledge of the event-probabilities should affect recency effects only in this case. It is suggested that when events are randomly sampled without replacement negative recency should occur to an increasing extent when subjects are aware

of the event-probabilities and when they are not aware of them, recency effects should change in the positive-negative direction. When events are randomly sampled with replacement no recency effects should occur with or without knowledge of the event-probabilities.

Method

The design was a 2 × 2 factorial, with randomized groups and 10 subjects in each group.

The dependent variable was a measure of recency which will be described in the next section, and the two factors were event replacement and information of event probabilities. Thus for two groups events were replaced and for two they were not replaced; and two groups were informed of the event probabilities, and two were not.

The experimental design and the predictions are summarized in Table I.

TABLE I

	Events replaced (*R*)	*Events not replaced* (*NR*)
Informed (I).. ..	Prediction: No recency effects	Prediction: Increasing negative recency
Not informed (NI) ..	Prediction: No recency effects	Predictions Change of recency effects in positive-negative direction

For convenience, the four groups will be referred to as Group (I,R), Group (I,NR), Group (NI,R) and Group (NI,NR).

All subjects were confronted with a deck of playing cards in which the ratio of red cards to black was either 2:1 or 1:2. (The ratio was reversed for half the subjects in each group in order to control for colour preference.) On each trial the subject took a card, predicting whether it would be red or black. All subjects received 200 trials.

Groups (I,NR) and (NI,NR) were presented with a deck of 200 playing cards and were required to turn over one card at a time until the deck was exhausted. After each trial the card was placed face down on the table.

Subjects in Group (I,NR) were informed of the ratio of red to black cards and were encouraged to glance through the deck in order to satisfy themselves of it. However, they were discouraged from sorting or counting the cards. When they announced that they were satisfied, they were asked to shuffle the cards, which were then cut by the experimenter and the experiment commenced.

Subjects in Group (NI,NR) were not informed of the ratio of red to black, but simply shuffled the cards which were cut by the experimenter.

Groups (I,R) and NI,R) were presented with a deck of 30 playing cards. They simply shuffled and cut 200 times, the cut card being replaced in the deck each time.

Subjects in Group (I,R) were informed of the ratio of red to black and encouraged to check it by counting. Subjects in Group (NI,R) were told nothing about the ratio of red to black.

Apparatus

A number of identical decks of playing cards were used from which the picture cards were removed. From these, four separate decks were compiled, two containing 200 cards and two containing 30 cards. In the decks of 200 the ratio of red to black cards was as near as possible 2:1 and 1:2 respectively, and in the decks of 30 the ratios were exactly 2:1 and 1:2.

Procedure

Subjects were run individually. Each subject sat at a table opposite the experimenter, who recorded the colour of the cut cards and the subject's responses.

The instructions read to each group were as follows:

Group (I,NR)

(*a*) This is a guessing experiment. Here is a stack of playing cards. I want you to shuffle it and then, on each trial turn over a card guessing beforehand whether the card you turn over will be a red one or a black one. After turning over each card place it face down before turning over the next.

(*b*) The stack contains twice as many red (black) as black (red) cards and I want you to satisfy yourself of this before we start. Examine the cards until you are satisfied that there are twice as many red (black). *I want you to genuinely guess on every trial.*

Group (NI,NR)

(*a*) As for Group (I,NR).

(*b*) There may or may not be the same number of red and black cards in the stack, but in any case *I want you to genuinely guess on every trial.*

Group (I,R)

(*a*) This is a guessing experiment. Here is an incomplete deck of playing cards. cards. It contains twice as many red (black) cards as black (red). Examine the cards. You see there are 30 altogether, 20 red (black) and 10 black (red). Count them to make sure.

(*b*) On each trial I want you to shuffle the cards and then cut a card, guessing between shuffling and cutting whether the card you cut will be a red one or a black one. *I want you to genuinely guess on every trial.*

Group (NI,R)

(*a*) This is guessing experiment. Here is an incomplete deck of playing cards. On each trial I want you to shuffle it and then cut a card, guessing between shuffling and cutting whether the card you cut will be a red one or a black one.

(*b*) As for Group (NI,NR) substituting "deck" for "stack."

The instruction in italics was read with particular emphasis. It was designed to discourage subjects from adopting an overall strategy (e.g. choosing one colour all the time). This seemed a necessary precaution, since for one group at least—(Group I,R)—the optimal strategy would clearly be to predict the more frequent event throughout, and such a strategy would preclude the occurrence of recency effects.

The decks of 200 cards used for Groups (I,NR) and (NI,NR) were thoroughly shuffled by the experimenter prior to the running of each subject. This is why subjects in Group (I,NR) who occasionally started to sort the cards while checking through them, were requested by the experimenter not to do so.

Subjects

Forty subjects were used, 10 in each group. For five subjects in each group the ratio of red to black cards was 2 : 1 and for the other five it was 1 : 2.

The subjects were all students from the Psychology Department at University College, London.

Results and Discussion

Recency effects are usually represented by means of a group curve in which percentage of predictions is plotted against length of run. This method of representation has the following disadvantages: (1) it obscures individual differences; (2) it does not provide a way of measuring recency effects; and (3) each point on the curve is based on a successively smaller number of observations (the length of a run is inversely related to its likelihood of occurrence), and hence, on a successively more unreliable estimate.

An alternative method of representing recency effects would be to classify all runs as either "short" or "long" and calculate a score for each subject based on his percentage of predictions following each kind of run. For example, suppose

a subject predicts an event 70 per cent. of the time following short runs and 40 per cent. of the time following long runs. By subtracting 70 from 40 one obtains a score of −30. The sign of the score indicates that the subject has displayed negative recency. Of course, this is a crude method since it only distinguishes two kinds of run, but it does provide a way of assigning a recency score to individual subjects and thus enables recency effects to be measured.

The recency scores presented below are calculated by comparing runs of 1, 2 and 3 against runs of 4 or more for the more frequent event, and runs of 1 against run of 2 and 3 or more for the less frequent event.

Before presenting the results the predictions outlined earlier will be summarized briefly:

Group (I,NR).	Negative recency throughout, increasing.
Group (NI,NR).	Recency effects should go in the positive-negative direction.
Group (I,R).	No recency effects.
Group (I,NR).	No recency effects.

Mean recency scores on the more frequent event are given in Table II.

TABLE II

	Recency scores		*Recency scores trials 1–100*	*Recency scores trials 101–200*
	Mean	*SD*	*Mean*	*Mean*
Group (I,NR)	−18·00	9·96	−20·80	−16·20
Group (NI,NR) ..	−25·60	26·12	−26·00	−25·20
Group (I,R)	3·60	8·04	8·30	−0·90
Group (NI,R)	−7·70	24·85	−4·20	−11·00

Table II shows that the predicted changes for Groups (I,NR) and (NI,NR) did not occur. For both groups the mean recency scores were slightly greater over the second hundred trials than over the first hundred. However, the other predictions were confirmed. An analysis of variance was carried out which showed a significant difference between "replacement" groups, but not between "information" groups, and no significant interaction. The results are given in Table III.

TABLE III

Source of variation	*Sum of squares*	*d.f.*	*Var. est.*	*F*
Between information levels ..	893·00	1	893·00	2·20, N.S.
Between replacement levels ..	3901·00	1	3901·00	9·59, $p < 0·01$
Information × replacement ..	34·23	1	34·23	—
Residual	14636·54	36	406·57	—
Total	19464·77	39	—	—

A Bartlett test showed significant nonhomogeneity of variance between the groups (B = 16·28, 3 *d.f.*, $p < 0·01$) and because of this, in calculating t ratios to test the recency predictions for Groups (I,NR), (I,R) and (NI,R), the standard error in each case was based on the group variance instead of on the residual variance.

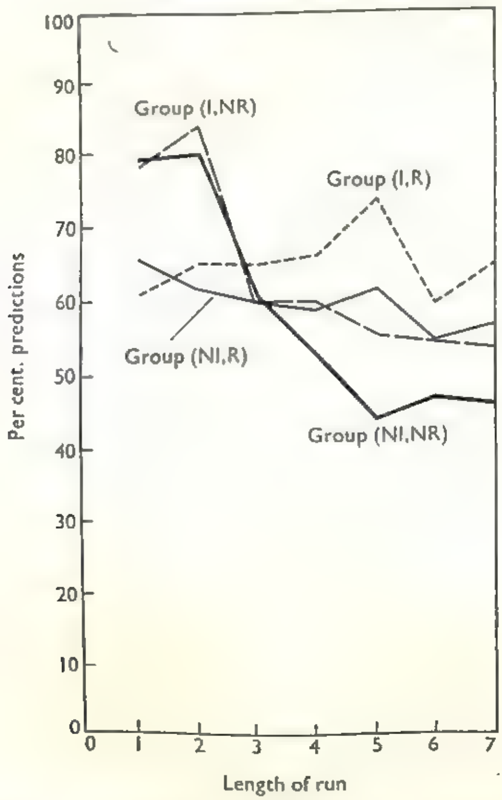
For Groups (I,NR), (I,R) and (NI,R) three t tests were carried out against a null hypothesis mean of zero. For Group (I,NR) the value of t was 5·42, giving a

one-tailed $p < 0{\cdot}0005$ (9 *d.f.*), and for the other two groups the value of t was less than unity. Thus the prediction that Group (I,NR) would display negative recency and that Groups (I,R) and (NI,R) would display no recency effects was confirmed. In addition, it may be noted that all subjects in Group (I,NR) had negative recency scores whereas for both other groups half of the subjects in each group had negative scores and half had positive scores.

No prediction was made concerning the overall recency effects for Group (NI,NR). It was only predicted that the recency effects would change in the positive-negative direction. The change did not occur, but this group displayed a high degree of negative recency throughout. Thus it would appear that regardless of whether subjects are informed of the event probabilities, when events are replaced they display no recency effects and when events are not replaced they display negative recency.

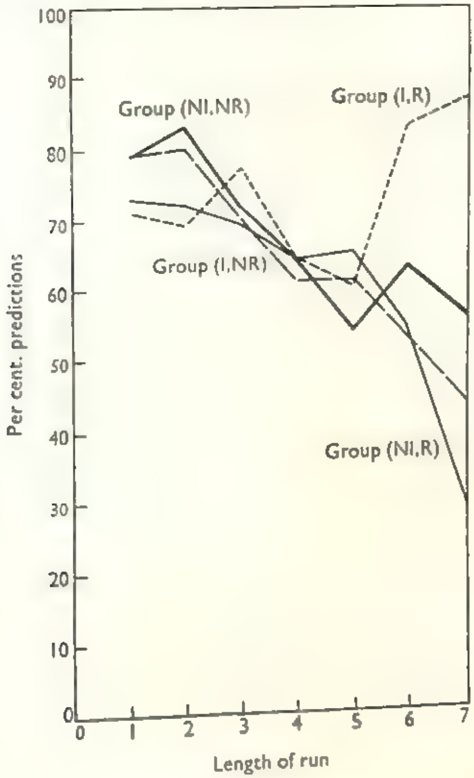
In Figures 1 and 2 the recency effects are represented in the usual way, group curves plotted for runs of 1 to 7 (runs of 8 and more are omitted owing to their relative infrequency of occurrence). It is seen that for Groups 1 and 2 the curve rises slightly before dropping. This seems to be a fairly common finding. Restle (1961) in comparing predictions from his "schema" theory with results from three experiments in which negative recency occurred, noted that in all three experiments the probability of an event being predicted was greater following a run of 2 than following a run of 1, although it decreased consistently for run lengths greater than 2 (Jarvik, 1951; Bush and Morlock, 1959; Goodnow, Rubenstein and Lubin, 1960). There seems to be no obvious explanation for this finding.

FIGURE 1



Percentage of prediction of more frequent event as function of length of run. Trials 1–100.

FIGURE 2



Percentage of prediction of more frequent event as function of length of run. Trials 101–200.

It is also seen (Fig. 2) that for Group (I,R), over the second 100 trials, the response probability is highest following runs of 6 and 7. This is a curious finding, but is probably attributable to chance, since these values are based on a relatively small number of cases.

In experiments where one event occurs more often than the other the author seldom discusses recency effects in relation to the less frequent event. Presumably this is because the less frequent event tends to occur in short runs only; for example, in this experiment runs of 4 or more on the less frequent event were very uncommon. Thus, if recency effects are defined as a gradual increase or decrease in response probability as a function of length of run, it is necessary to have more than three run lengths in order to represent them satisfactorily.

In this experiment, recency effects on the less frequent event were measured by comparing response probabilities after runs of 1 with those after runs of 2 and 3. The mean recency scores for the four groups were:

Group (I,NR), 1·00;
Group (I,R), −5·60;
Group (NI,NR), −7·30;
Group (NI,R), −3·20.

It would appear from this that none of the groups displayed negative recency on the less frequent event. However, it will be remembered that the recency scores were calculated in a different way from those on the more frequent event. Had the recency scores on the more frequent event been based only on runs of 1, 2 and 3 it is possible that no recency effects would have been apparent. In fact, this is quite

likely, since the groups which displayed negative recency, Groups (I,NR) and (NI,NR), showed a stronger tendency to predict the more frequent event after runs of 2 than after runs of 1.

When the recency scores on the more frequent event were calculated in the same way as those on the less frequent event, by excluding runs of 4 or more and basing the scores on a comparison of the response probabilities after runs of 1 with those after runs of 2 and 3, the mean recency scores were:

Group (I,NR), −3·20;	Group (NI,NR), −2·70;
Group (I,R), 2·70;	Group (NI,R), −4·20

Clearly, when recency scores are calculated in this way there is no evidence of the negative recency which the previous analysis yielded. It follows from this that one can assert nothing about recency effects on the less frequent event, because it only occurred in short runs. Had it occurred in longer runs similar recency effects to those on the more frequent event might have been observed. To achieve this would have required a much longer series of trials, or else a rigged one.

In general, several of the predictions were not supported, and since these were based on purely rational considerations it seems reasonable to conclude that subjects do not behave entirely rationally in the two-choice situation. However, they do not seem to behave as irrationally as some writers have claimed, inasmuch as they do not display negative recency in anticipating events in a series which they can see is random. Belief in some kind of event-dependence would appear to be a necessary condition for the occurrence of negative recency, although whether this covers any kind of event-dependence remains to be seen. All that has been shown here is that subjects display negative recency when confronted with a particular kind of event-dependence; moreover, one which makes it rational to display negative recency, and that when events are clearly independent they display no recency effects.

This research was carried out at University College London as part of a Ph.D. investigation, written in 1962. The author is indebted to Dr. A. R. Jonckheere for his advice and criticism in connection with the research.

References

Bush, R. R., and Morlock, H. C. (1959). Test of a general conditioning axiom for human two-choice experiments. *Univ. Pennsylvania Psychol. Lab. Memo MP-1.*

Cohen, J. (1960). *Chance, Skill and Luck*, 29. London.

Goodnow, J. J., Rubenstein, I., and Lubin, A. (1960). Response to changing patterns of events. *Amer. J. Psychol.*, **73**, 56–67.

Hyman, R., and Jenkins, N. S. (1956). Involvement and set as determinants of behavioural stereotypy. *Psychol. Rep.*, **2** (Monogr. Suppl. No. 3).

Jarvik, M. E. (1951). Probability learning and a negative recency effect in the serial anticipation of alternative symbols. *J. exp. Psychol.*, **41**, 291–7.

Morse, E. B., and Runquist, W. N. (1960). Probability-matching with an unscheduled random sequence. *Amer. J. Psychol.*, **73**, 603–7.

Restle, F. (1961). *Psychology of Judgement and Choice.* New York: Wiley, 117–21.

Manuscript received 12th June, 1965.

SHORT-TERM RECOGNITION MEMORY FOR SINGLE LETTERS AND PHONEMIC SIMILARITY OF RETROACTIVE INTERFERENCE

BY

WAYNE A. WICKELGREN

From the Department of Psychology, Massachusetts Institute of Technology

Copying 12 letters produces more retroactive interference in recognition memory for a single letter when the interference letters possess a vowel sound in common with the letter to be remembered than when they do not. Compared to interference lists that do not include the presented letter, inclusion in the interference list of the letter to be remembered improves recognition memory when the other interference letters have no vowel sound in common with the letter to be remembered, but not otherwise. False recognition rates are greater when the test letter contains a vowel sound in common with the presented letter than when the vowel sounds of these two letters are different. The findings are in complete accord with analogous findings for short-term recall and indicate that short-term recognition memory uses the same phonemic-associative memory system as short-term recall.

Introduction

Recent evidence suggests that letters, digits and words are coded in short-term memory, not as atomic units, but as combinations of vowel and consonant phonemes. Conrad (1964) established that letters whose pronunciation ends in "ē" (B, C, P, T, V) tend to be confused with each other in recall and letters whose pronunciation begins with "ĕ" (F, M, N, S, X) tend to be confused with each other in recall. This was true even though the letters were presented visually at a rate of 0·75 sec./letter under conditions where the probability of perceptual error was known to be negligible. Wickelgren (1965*b*) replicated these findings for "ē" and "ĕ" letters under somewhat different conditions and found the same relationship between acoustic similarity and intrusion errors for other vowel and consonant similarities. The relationship between acoustic similarity and intrusions occurred not only in intrusions of letters for letters, but also in intrusions of letters for digits and digits for letters.

A second line of evidence in support of the functional importance of phonemic coding in short-term memory is a study by Conrad and Hull (1964) which found that lists of letters with a common vowel sound were more difficult to recall than lists of letters with different vowel sounds. Similar results were obtained with lists of words of similar versus dissimilar pronunciation (Conrad, 1963).

The functional significance of phonemic units in short-term memory is further supported by the finding that proactive and retroactive interference in short-term recall are affected by the acoustic similarity of the interfering material. Wickelgren (1965*a*, 1966) found that an interference list containing no letters in common with the correct letter(s) produced greater interference if the interference letters had the same vowel sound ("ē" or "ĕ") as the correct letter(s) than if the interference letters had a vowel sound different from the correct letter(s). The findings of these studies were interpreted in support of the hypothesis that there is only one internal representative of any particular phoneme in short-term memory, regardless of the number of times that phoneme is presented in the same or other letters. Thus, there is only one internal representative of "ē" and only one internal representative of "ĕ."

When any "ē" letter (B, C, D, E, G, etc.) is presented, this one internal representative of "ē" is activated, along with the internal representative of whatever consonant accompanied the vowel "ē."* According to this theory it is clear why an interference list composed of "ē" letters should produce more retroactive interference in the recall of other "ē" letters than an interference list composed of "ĕ" letters, and vice versa. If the interference list activates the same vowel representative as the original list, then associations will be formed between this vowel representative and consonant representatives in both the original and interference lists, and these associations will compete with each other. If the interference list activates other vowel representatives, this type of associative interference will not occur.

It seems difficult to account for the present findings and previous findings on phonemic similarity without assuming that items are coded in short-term memory as sequences of phonemes, at some level of analysis. Naturally, we cannot say from these data whether the phoneme is the ultimate unit of coding in short-term memory. The results of Miller (1956) and Miller and Nicely (1955) indicate that in auditory recognition of vowels and consonants there is a level of analysis beyond the phonemic level, namely distinctive feature analysis. It may be that the distinctive feature is also a more basic unit of coding in short-term memory, and perhaps there are levels of analysis even more basic than distinctive features. The present findings indicate only that, whatever the most basic units are, these basic units combine to represent phonemes, which in turn combine to represent letters, digits, words, etc.

All of the previous studies of acoustic similarity and short-term memory have studied recall. The purpose of the present study is to determine if analogous effects of acoustic similarity can be established for short-term recognition memory. This is motivated by two hypotheses: first, the phonemic coding hypothesis just discussed and second, the hypothesis that recall and recognition are two different ways of using the same memory system.

Although it may seem natural to assume that the trace used in recognition is no different than the trace used in recall, this is not a logical necessity. There are at least two ways in which the underlying trace might differ for recognition and recall. First, if a subject knows prior to learning what the retention test will be, he may store items differently. Second, whether or not he knows the retention test in advance, traces may always be established in several different memory systems, one of which is used in recognition and another of which is used in recall.

Existing evidence supports the hypothesis that recognition and recall use the same memory system. Number of presentations, proactive interference and retroactive interference affect recognition memory in the same way that they affect recall (Egan, 1958; Peixotto, 1947; Postman, 1952; Florès, 1960; Shepard and Teghtsoonian, 1961). However, there are several, rather different, procedures for studying recognition, and one should be cautious in generalizing results from one procedure to another. It may make some difference whether one is asked: "Did you ever see this item before?", "Did you see this item in that list you saw at such and such a time?", "Which one of these n items was in that list you saw at such and such a time?". "Which ones of these n items were in that list you saw at such and such a time?", "Which of these two items did you see more recently?", etc.

The present experiment on recognition memory used the question, "Did you see this item in that list you saw before the interference list?", and the list before the interference list consisted of just one letter. In addition, subjects were required to indicate the confidence they had in their "yes-no" decision on a five-point rating

* Complex syllabic nuclei, such as ē (|iy|), will be called phonemes in this paper.

scale. This procedure enables one to compute nine points of a "memory operating characteristic" (MOC curve) for each recognition condition and thus control for differences in false recognition rate when studying correct recognition rate. Rather extensive discussions of the use of the operating characteristic in studies of recognition memory are to be found in Egan (1958) and Norman and Wickelgren (1965).

Method

Procedure

The procedure for each trial was as follows: One sec. after a ready signal, a single letter was presented, followed after 0·5 sec. by a tone and then the interference list of 12 letters presented at the rate of 0·5 sec./letter. The interference list was followed by a tone and, after 0·5 sec., by a single letter, to which subjects were to respond "yes," if they thought the test letter was identical to the originally presented letter, and "no," if they thought the test letter was different from the originally presented letter. Subjects were also instructed to indicate confidence in their decision on a five-point scale, where "5" meant "most confidence" and "1" meant "least confidence." The presented letter, the 12 interference letters, and the test letter were to be copied as they were being presented. The presented letter was to be covered immediately after being copied. Subjects were given 10 sec. in which to make their decision and indicate their confidence, so a trial required about 20 sec., followed immediately by the next trial. The entire experiment was recorded on tape and lasted approximately 50 min.

Design

There were 36 conditions in the experiment. Two types of presented letters were used, "ē" letters (B, C, D, E, G, P, T, V, Z-pronounced "zee") and "ĕ" letters (F, L, M, N, S, X). Three types of test letters were used: (1) the presented letter, (2) a different letter, from the same acoustic class as the presented letter, and (3) a different letter, from the other acoustic class (different class). Six types of interference letters were used: (1) same acoustic class as presented letter, presented letter not included; (2) different acoustic class from presented letter, presented letter not included; (3) both acoustic classes (six of the interference letters from the "ē" class and six from the "ĕ" class), presented letter not included; (4) same acoustic class as presented letter, presented letter included at least once; (5) different acoustic class from presented letter, except that the presented letter was included at least once; (6) both acoustic classes, presented letter included at least once. Subject to the above restrictions, the interference sequences were constructed by drawing randomly with replacement from the populations indicated in the definition of the conditions.

Subjects were 38 M.I.T. undergraduates taking psychology courses who participated in the experiment as part of their course requirements. Subjects were run in three approximately equal groups; every subject being tested in all 36 conditions. Conditions were randomly ordered in blocks of 36 trials, and there were four blocks in the experiment, for a total of 144 trials.

Analysis

Let us consider a rather general two-stage model of the recognition process in which both a memory and a decision system operate to determine the response. All items, presented or not presented, are assumed to be characterized by a real-valued strength in memory. The decision system maps these strengths on to responses by comparing the strength of the test item with some criterion value. Items that exceed the criterion receive a response of "yes," otherwise they are assumed to be non-presented items and receive a "no" response. According to this reasoning, false recognitions contain valuable information. By forcing the subject to vary his bias (and thereby his criterion strength) while holding the presentation conditions constant (and, hopefully, holding constant his memory of the items), one can trace out the relative strengths of the distributions of presented and non-presented items. The relation between correct and false recognition rates will be referred to as the memory operating characteristic (MOC curve), which is directly analogous to the receiver operating characteristic (ROC curve) in signal detection theory.

In the binary choice experiment just described each point must be obtained in what is essentially a separate experiment in which the subject's bias to say "yes" is manipulated

by changing his instructions or pay-offs. Fortunately, a more economical technique has been developed (Egan, Schulman and Greenberg, 1959; Pollack and Decker, 1958). After making his binary decision a subject can indicate his confidence in that decision on a rating scale. We interpret his confidence as a direct reflection of the strength of the item along a unidimensional scale from a "most confident yes" to a "most confident no." This permits us to get several points of the operating characteristic in one experiment.

To each test item a subject must respond with one of ten decision-confidence pairs. Let $i = 1, 2, \ldots, 5, 6, \ldots, 10$ represent "yes" with confidence "5" (greatest confidence), "yes" with confidence "4," . . ., "yes" with confidence "1" (least confidence), "no" with confidence "1," . . , "no" with confidence "5." Let $f_i(x)$ represent the total frequency (over all four blocks and all 38 subjects) with which response i occurred in condition x. Let $r_i(x) = f_i(x)/\sum_{i=1}^{10} f_i(x)$ represent the relative frequency with which response i occurred in condition x. Let $R_i(x) = \sum_{j=1}^{i} r_j(x)$ represent the cumulative relative frequency with which responses 1 through i occurred in condition x. The MOC curve is a plot of $R_i(x)$, the (correct) recognition rate for some condition in which the test item is the same as the presented item, against $R_i(y)$, the (false) recognition rate for some condition in which the test item is different from the presented item.

In common sense terms, it is relatively meaningless to compare conditions with respect to correct recognition rate (correct "yes" responses) unless you also compare the conditions with respect to false recognition rate (incorrect "yes" responses). If one can obtain correct recognition rates for several different values of the false recognition rate, then one can plot an MOC curve for a condition and compare that curve to the MOC curve for another condition. If one curve lies above the other curve, then recognition memory is better in the condition with the higher MOC curve. The confidence judgement technique allows one to determine one point on the MOC curve for each possible cut-off along the decision-confidence continuum. In this experiment there are ten decision-confidence pairs, hence nine cut-offs and nine points on the MOC curve for each condition.

Results

Similarity of interference letters

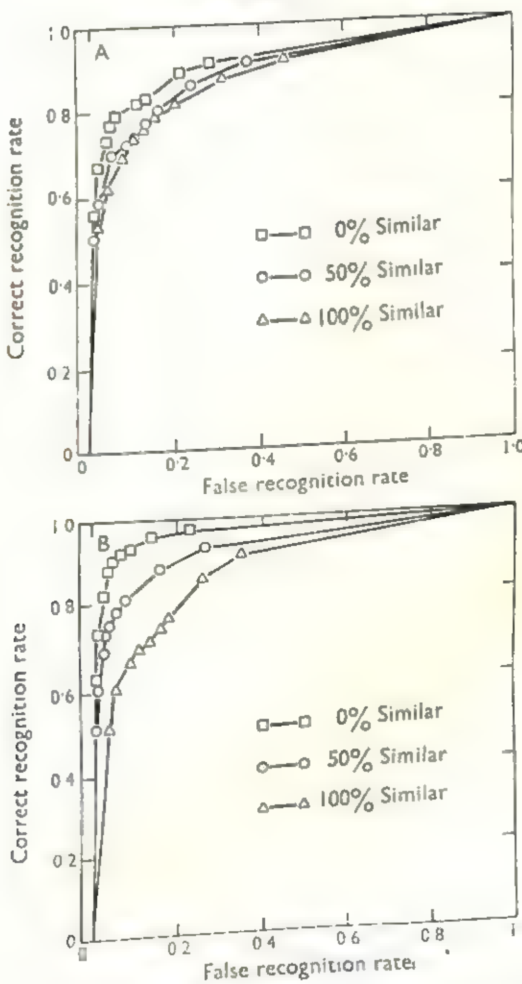
The most appropriate estimate of the false recognition rate for each of the interference conditions comes from the condition in which the test item is different from the presented item, but from the same acoustic class. Using the condition in which the test item is from a different acoustic class would be confounded by possible differences in the false recognition rate for different classes of letters. Therefore, throughout this section "false recognition rate" will mean the "false recognition rate for the condition in which the test item is different from the presented item, but from the same acoustic class."

Figures 1A and 1B plot the correct recognition rate against the false recognition rate (MOC curve) for each of the six interference conditions averaged over the two presented letter conditions. Figure 1A contains the MOC curves for the three conditions in which the presented letter never occurred in the interference list; Figure 1B contains the MOC curves for the three conditions in which the presented letter occurred at least once in the interference list. In both Figures 1A and 1B recognition memory is superior when the retroactive interference is composed of acoustically different letters than when it is composed of acoustically similar letters. In both cases the condition with half similar and half different letters is intermediate in recognition memory performance. Retroactive interference in short-term recognition memory is clearly a function of the acoustic similarity of the interfering materials, the relationship being greater interference with more similar material.

Unfortunately, no statistical test has been developed to determine if two operating characteristics are significantly different. However, it is possible to give an approximate idea of the reliability of the difference between two operating characteristics by applying the Kolmogorov-Smirnov two-sample test to the correct recognition

rates at equal false recognition rates. This is not completely valid since it does not take into account the error in estimating the false recognition rates. Fortunately, all the paired comparisons are in the same direction, and many of them are highly significant by this test. In Figure 1A, the 0 per cent. similar condition has a significantly greater correct recognition rate than the 100 per cent. similar condition at equal false recognition rates ($D = 0{\cdot}14$, $p < 0{\cdot}01$, two-tailed), but the other two paired comparisons are not significant. In Figure 1B, the 0 per cent. similar condition is significantly superior to the 50 per cent. condition ($D = 0{\cdot}13$, $p < 0{\cdot}01$, two-tailed), the 50 per cent. condition is significantly superior to the 100 per cent. similar condition ($D = 0{\cdot}24$, $p < 0{\cdot}001$, two-tailed), and of course, the 0 per cent. versus 100 per cent. comparison is significant at well beyond the 0·001 level.

FIGURE 1



MOC curves for single letters after copying 12 interference letters that were 0 per cent., 50 per cent. or 100 per cent. similar to the letter to be remembered. In A the letter-to-be-remembered never appeared in the interference list. In B the letter-to-be-remembered appeared at least once in the interference list and was counted as a "similar" letter. "Similar" letters had a vowel phoneme (ē or ĕ) in common with the letter-to-be-remembered; "different" letters had no phoneme in common with the letter-to-be-remembered.

Comparing the curves in Figure 1A with those in Figure 1B demonstrates that including the presented letter in an interference list composed of acoustically different letters improves later recognition memory, but including the presented letter in an interference list composed of acoustically similar letters is of little or no benefit to later recognition memory. The mixed interference is again intermediate. These results may be explained as follows. First, the occurrence of the presented item in the interference list very likely aids later recognition memory only when subjects recognize it as the presented letter at the time of its occurrence in the interference list. Second, recognition of the presented letter is more likely after different letters than after similar letters, as already established. Viewed in this manner, the comparison of Figures 1A and 1B provides further support for the hypothesis that retroactive interference in recognition memory is a direct function of acoustic similarity.

Similarity of test letter

Table I presents the cumulative false recognition rates for the similar and different test letter conditions, averaged over the two presented letter conditions and the six interference conditions. A chi-square test on the frequency of "yes" versus "no" responses in the two false recognition conditions is significant at the 0·01 level ($\chi^2 = 9{\cdot}27$). Since all letters are equally probable in the two conditions, when

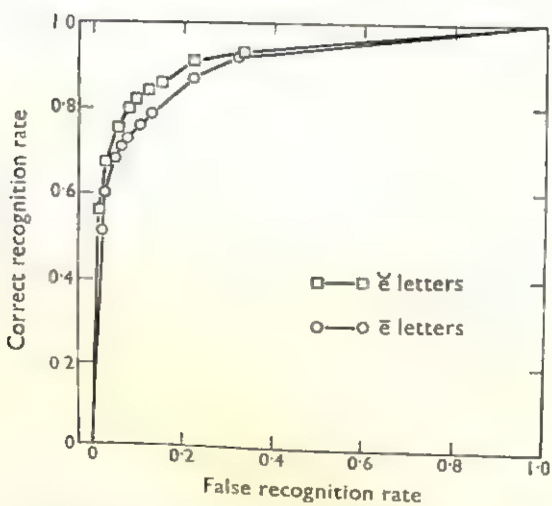
TABLE I

CUMULATIVE FALSE RECOGNITION RATES FOR SIMILAR AND DIFFERENT TEST LETTERS

($R_i(x)$ in percentage)

	Y5	*Y4*	*Y3*	*Y2*	*Y1*	*N1*	*N2*	*N3*	*N4*	*N5*	N
Similar	1·6	2·8	5·2	7·0	8·6	11·3	14·1	22·0	32·8	100	1740
Different	0·6	1·7	3·2	4·4	5·9	7·8	11·5	18·6	28·9	100	1793

FIGURE 2



MOC curves for "ē" letters and "ĕ" letters.

averaged over all values of the other two variables in the experiment, we cannot attribute this result to intrinsic differences in false recognition rate for different letter classes. Thus we conclude that a test letter possessing a vowel phoneme in common with the presented letter is more likely to be falsely recognized than a test letter possessing no phoneme in common with the presented letter.

Nature of the presented letter

Figure 2 presents the MOC curves for "ē" letters and "ĕ" letters, averaged over the six interference letter conditions and using test letters from the same acoustic class to estimate the false recognition rate. Recognition memory for "ĕ" letters is clearly better than recognition memory for "ē" letters. This is in line with the finding of Wickelgren (1965*a*) that free recall is better for "ĕ" letters than for "ē" letters. This result might be due to the size of the confusion class or countless other factors, and nothing more definite can be said.

Discussion

In recall, the memory system, acting on whatever cues are given, must produce the test item. In recognition, the memory system, acting on the test item, must produce a "yes-no" decision. The stimulus conditions of the original presentation can be identical in the two cases. The stimulus conditions at the time of retrieval are obviously different in the two cases. The question presently at issue is whether the same memory system is used to produce above-chance performance in both cases. To the extent that recall and recognition memory are affected in the same manner by the same variables, it is parsimonious to assume that both use the same system.

The present experiment establishes that the acoustic similarity of retroactive interference affects short-term recognition memory in the same way that it affects short-term recall, namely, greater interference from similar material than from different material.

Finding that an acoustically similar test item produces more false recognitions than an acoustically different test item is analogous to the finding in recall experiments that intrusions tend to be acoustically similar to the correct item (Conrad, 1964; Wickelgren, 1965*b*). One could say for both recall and recognition memory that false positives tend to be acoustically similar to the correct item.

Finally, "ĕ" letters are remembered better than "ē" letters under either the recall or recognition testing procedures. Every possible comparison between the results of this experiment and previous experiments on recall supports the hypothesis that short-term recall and recognition use the same memory system.

This work was supported primarily by grant, MH 08890–01, from the National Institute of Mental Health, U.S. Public Health Service. Further aid was received from grant, MH 05673–03, to Hans-Lukas Teuber.

References

Conrad, R. (1963). Acoustic confusions and memory span for words. *Nature*, **197**, 1029–30.

Conrad, R. (1964). Acoustic confusions in immediate memory. *Brit. J. Psychol.*, **55**, 75–84.

Conrad, R., and Hull, A. J. (1964). Information, acoustic confusion and memory span. *Brit. J. Psychol.*, **55**, 429–32.

Egan, J. P. (1958). *Recognition Memory and the Operating Characteristic*. Indiana University, Hearing and Communication Laboratory—AFCRC–TN–58–51, AD–152650, 1958.

EGAN, J. P., SCHULMAN, A. I., and GREENBERG, G. Z. (1959). Operating characteristics determined by binary decisions and by ratings. *J. acoust. Soc. Amer.*, **31**, 768–73.

FLORÈS, C. (1960). Étude de l'évolution des interférences a l'intrieur des épreuves de reconnaissance. *Année Psychol.*, **60**, 339–48.

MILLER, G. A. (1956). The perception of speech. In *For Roman Jakobson.* Hague, 353–9.

MILLER, G. A., and NICELY, P. E. (1955). An analysis of perceptual confusions among some English consonants. *J. acoust. Soc. Amer.*, **27**, 338–52.

NORMAN, D. A., and WICKELGREN, W. A. (1965). Short-term recognition memory for single digits and pairs of digits. *J. exp. Psychol.*, **70**, 479–89.

PEIXOTTO, H. E. (1947). Proactive inhibition in the recognition of nonsense syllables. *J. exp. Psychol.*, **37**, 81–91.

POLLACK, I., and DECKER, L. R. (1958). Confidence ratings, message reception, and the receiver operating characteristic. *J. acoust. Soc. Amer.*, **30**, 286–92.

POSTMAN, L. (1952). Retroactive inhibition in recall and recognition. *J. exp. Psychol.*, **44**, 165–9.

SHEPARD, R. N., and TEGHTSOONIAN, M. (1961). Retention of information under conditions approaching a steady state. *J. exp. Psychol.*, **62**, 302–9.

WICKELGREN, W. A. (1965*a*). Acoustic similarity and retroactive interference in short-term memory. *J. verb. Learn. verb. Behav.*, **4**, 53–61.

WICKELGREN, W. A. (1965*b*). Acoustic similarity and intrusion errors in short-term memory. *J. exp. Psychol.*, **70**, in press.

WICKELGREN, W. A. (1966). Phonemic similarity and interference in short-term memory for single letters. *J. exp. Psychol.*, in press.

Manuscript received 29th May, 1965.

SUSCEPTIBILITY TO THE MÜLLER-LYER ILLUSION AND ITS RELATIONSHIP TO DIFFERENCES IN SIZE CONSTANCY

BY

VERNON HAMILTON

From the Department of Psychology, University of Reading

The hypothesis that susceptibility to the Müller-Lyer illusion is the result of normal constancy scaling, misapplied, was submitted to direct test. No significant correlations between illusion error and size constancy estimates were obtained. Also invalidated were hypotheses that under-constancy is correlated with non-susceptibility to the illusion, and that over-constancy is correlated with greater illusion error. The results suggest that an approach to the explanation of illusion effects by means of analysing individual differences in size constancy, in intelligence and preferred "perceptual style," might be fruitful. Some tentative suggestions are made concerning the role of perceptual inference, abstraction and analysing.

Introduction

The hypothesis that the optical illusions are a special case of processes which also operate in size constancy judgements has recently received renewed attention (Tausch, 1954, 1955; Gregory, 1963, 1964*a*, 1964*b*). Gregory's statement of this relationship with which this study has been specifically concerned is that "The illusion figures may be thought of as flat projections of typical views of objects lying in three-dimensional space . . .," and, that those parts of the illusion figures ". . . corresponding to distant objects are expanded and the parts corresponding to nearer objects are reduced" (1963, p. 678). He infers that perceptual behaviour in relation to the illusions is thus similar if not identical to that operating in the estimation of size over distance, and that the two parts of the Müller-Lyer figure will in fact be perceived as, say, corners of a house or a room in the proximal and distal position, respectively. Gregory proceeded to develop an ingenious experimental condition in which the postulated close relationship between size and perceived depth of the illusion could be verified. He found a correlation of about 0·9 between the two parameters, and produced curves showing a very close match indeed between size of the illusion and perceived depth of the illusion figure (Gregory, 1965).

A partial explanation of the visual illusion in terms of constancy scaling was also given by Thurstone (1944) in his massive factorial exploration of perceptual processes, but this remained invalidated by the absence of significant correlations between any of the visual illusions and the index of size constancy. Apart from the medium-sized correlations between the different illusions, Thurstone's findings would seem to be noteworthy because of the correlations between the Müller-Lyer illusion on the one hand, and the Gottschaldt hidden figures and spatial intelligence on the other. The former finding forms part of the reasoning of Gardner (1961) in his attempts to account for susceptibility to the illusions in terms of field articulation and scanning. These processes are part of the system of cognitive control dimensions which have been postulated by investigators at the Menninger Foundation in attempts to define individual differences and intra-individual consistencies in perceptual operations.

Inquiries into individual differences in responses to illusion figures may be considered overdue. Since the phenomena possess considerable universality which have defied many previous attempts at explanation, *differences in degree* of susceptibility to the well-known situations may provide a new avenue of attack. In the past, individual differences have been frequently ignored, as have been individual differences in, e.g. reaction times, apparent movement and also the constancies. Gregory, too, has remained silent on the question of individual differences, an approach that is shared by Tausch (1954, 1955) and the Japanese investigators as summarized by Oyama (1960). The involvement of individual differences when considering a postulated relationship between size constancy and susceptibility to the illusions would seem to be particularly pertinent in view not only of Thouless's findings of various kinds of individual differences in perceptual constancy (1931*a*, 1931*b*, 1932), but because of more recent findings of individual and group differences in size constancy between normal subjects and subjects with abnormal types of behaviour (e.g. Raush, 1952; Crookes, 1957; Perry, 1961; Hamilton, 1963). While the notion of individual differences is inseparable from the predictions of a co-variant relationship between degree of illusion and degree of size constancy, a direct test of this hypothesis does not appear to have been carried out since Thurstone's study. Given Gregory's hypothesis it was felt that this kind of direct test was mandatory.

Method

Apparatus

Size constancy was assessed in relation to three standard objects, a white rod, a white square in the diagonal presentation and a standard sized playing card. The variable in the case of the square was a manually operated diamond, variable in size and adjusted by the experimenter according to instruction from subjects; for the other objects the variable was a white rod, operated by means of a handle by subjects. The size estimates with which we are concerned here were carried out under full cue conditions on a table painted matt black 3·10 m. long, 70 cm. wide and 79 cm. high. Distances between the subject and the standard and variable were 1 m. and 3 m. respectively, when the subject's head was resting in a chin rest. The horizontal angular separation between standard and variable was fixed at 5°, this being the smallest angular separation at which Joynson (1958*a*, 1958*b*) obtained close to constancy judgements. Because of the angular closeness of standard and variable, the variable was permanently raised giving a vertical angular separation of roughly 7°. The veridical, objective sizes of the standards were 15 cm., for diamond and rod, and 8·9 cm. for the playing card. Each subject made four judgements or settings for each object, starting twice from the minimal position of the variable and twice from the maximal position, with the starting positions alternating.

The Müller-Lyer apparatus was of the traditional sleeve and slide pattern. The central standard line was 7·5 cm. long, the arms of the angles were 2 cm., and the size of the fin angle was 60° for the standard and 300° for the variable, internal and external respectively. The maximum extent of the central variable line was 10 cm., and the maximum measurable error of over-estimation of length was 53 per cent. Each subject manually made eight "equality" settings of the variable, i.e. that part of the figure with one non-arrow-form end. For the starting position of one series of four adjustments the variable was extended to its maximum, while the starting point for the second series of four adjustments was from the point of minimal extension of the variable. In each series of adjustments the variable occupied successively the position of the four compass points in relation to the standard. No attempt was made on this occasion to control for order effect in the sequence of orientations and directions of adjustment. The constant sequence of orientations of the variable in relation to the standard was "East," "North," "West," "South," and minimal extension of the variable preceded maximal extension. All adjustments were made with the apparatus lying on a table in front of the subject at comfortable "reading" distance holding the standard end in one hand and the variable end in the other.

The data are expressed in terms of per cent. error, i.e. per cent. over-estimation of the variable with its large external fin angle in relation to the standard with its small

internal fin angle. Size constancy judgements were expressed simply in terms of raw size adjustments and per cent. error in relation to veridical standard size. Any form of conversion of the raw data was considered an unnecessary exercise which would add nothing of material value to the present findings.

Instructions

Considerable efforts were made to obtain a homogeneous attitude towards the size constancy estimates and to minimise the analytic or apparent size approach. Subjects were told that they were to carry out objective equality matchings so that standard and variable would be the same size after adjustment if they were put side by side or if they were measured. Since the subjects were children, some of them of low intelligence, they were given in the instruction the example of a ship out at sea and the same ship close to shore to demonstrate the general idea of the relationship between perceived size and distance. Instructions with variations on this theme were continued until there appeared to be a general grasp of the relationship where it had not been present before.

For the Müller-Lyer illusion the instructions for equality adjustments appeared to be unequivocal for the subjects, but several had to be asked not to make snap decisions, and to make their responses carefully and after reflection.

Subjects

A necessary requirement of the study was to have available an experimental group showing considerable variance in their size constancy responses because this would increase the probability of obtaining a significant correlation between the two variables. On the basis of considerations not pertinent to this report, and on the basis of other findings by the writer of individual differences in size constancy, it was predicted that children of low intelligence, attending a school for the educationally subnormal would not have achieved the degree of cognitive maturity which finds expression in full size constancy responses normally considered to be available in children by the age of 10 (e.g. Vernon, 1952). For this reason children from a secondary modern school as well as children from a school for the educationally subnormal were used as subjects. Since sex is known to interact with size constancy (e.g. Thouless, 1932; Crookes, 1957) only boys were used. Table I shows the basic statistics of the subjects.

TABLE I

	*I.Q.**			*Age*			N
	Mean	*S.D.*	*Range*	*Mean*	*S.D.*	*Range*	
Non-E.S.N. ..	95·2	9·07	82–112	12·2	0·35	11·5–12·8	20
E.S.N. ..	74·1	5·35	65–85	12·6	0·66	11·4–13·6	20

* See text for test data used.

A word is required concerning the validity of these I.Q.'s. Assessments for the specific purpose of this investigation were not practicable. The secondary modern school children were, however, in their first year in the secondary school, and the available Moray House Quotients from the so-called intelligence paper appeared to offer a sufficiently good approximation to what would have been obtained on another verbally loaded test. For the E.S.N. children the last individual intelligence test results in the school medical officer's records was somewhat reluctantly accepted. On subjective impression some of these were a little on the low side for some of the subjects. Basically, however, the samples were undoubtedly widely different in ability, and there is, of course, in addition the operational definition of this difference in their different scholastic placements, where for the secondary modern school boys, the low I.Q.'s were from subjects in the B and C streams of their year.

Hypothesis

The main purpose of the present experiment was to test out directly the postulated relationship between degree of illusion shown with the Müller-Lyer figure and degree of size constancy. Gregory has shown that with certain experimental conditions and instructions there exists a close co-variance between degree of the illusion and the perceived depth of the illusion figure. This is not a direct method for assessing the relationship between size constancy and susceptibility to a visual illusion because perceived depth is only one piece of information entering into size estimation and because recent studies (e.g. Rump, 1961) have cast doubt on the frequently assumed linear relationship between perceived size and perceived distance. It therefore seemed more appropriate to test Gregory's hypothesis by assessing the co-variance of illusion error and constancy scaling by obtaining as a measure for the latter actual data on the perception of *size* over distance, rather than data on the perceived degree of depth uncontaminated by illusion-inducing configurations. On the basis of Gregory's conceptualizations, degree of illusion and degree of constancy should be significantly correlated.

Certain corollaries follow from a postulated link between size constancy and illusion error. In the present context adequate constancy scaling would ideally lead to only minor errors in the perception of size over distance, and any corrections of visual angle or phenomenal percepts would be in the opposite direction to that suggested by them. This tendency may be described as over-constancy and evidence of its existence in adults has been variously obtained (e.g. Hamilton, 1963, 1965). It suggests the presence of compensatory mechanisms which are likely to operate in any condition in which size judgements at a distance are involved, and which, presumably operate, therefore, in conditions in which constancy scaling is applied to size judgements of illusion figures seen at a distance. If the illusion error with the Müller-Lyer figure is the result of normal constancy scaling, then over-correction of apparent size should equally apply to illusion and non-illusion figures. This, in the case of the illusion figure and on the basis of the constancy scaling hypothesis, should lead to a larger illusion error or greater susceptibility to the illusion. In a similar way it can be argued on the basis of Gregory's hypothesis that under-constancy in size estimates, i.e. inadequate constancy scaling should be accompanied by a reduction in illusion error.

On the basis of these considerations three hypotheses reflecting the substance of Gregory's theory are formulated:

(1) Degree of illusion and size constancy are significantly correlated.

(2) Over-constancy is found in association with greater susceptibility to the illusion.

(3) Under-constancy is found in association with reduced susceptibility to the illusion.

The following definition of terms are adopted: full constancy means that subjects will adjust the variable in the distal position to a size equalling or very closely approximating the size of the standard object in the proximal position; over-constancy for the purpose of the present discussion will refer to size estimates in which the variable in the distal position is set to veridical object size minus 5 per cent. or more of the standard in the proximal position; under-constancy will refer to size estimates in which the variable in the distal position is set to veridical object size plus 5 per cent. or more of the standard in the proximal position.

Results

Size estimates for the three standards employed in the study, and the average per cent. error of estimation on the Müller-Lyer figure for all 40 subjects are tabulated in Table II. They are ranked from the subject with the least error on, or susceptibility to the illusion, to the subject with the greatest illusion error.

TABLE II

Müller-Lyer and Constancy Estimates (in cm.)

Subject No.	*Per cent. error Müller-Lyer*	*Diamond*	*Rod*	*Playing card*	*Subject No.*	*Per cent. error Müller-Lyer*	*Diamond*	*Rod*	*Playing card*
17	15·163	16·175	16·025	10·875	34	26·660	19·025	14·750	9·250
37	16·663	17·150	17·850	16·725	11	27·826	15·475	12·325	9·250
18	18·329	14·325	14·875	10·075	32	28·326	16·525	12·325	10·950
22	19·662	16·950	15·175	9·375	39	29·159	15·225	15·100	12·375
14	19·828	11·250	16·450	8·200	19	29·493	15·575	14·225	8·500
15	19·828	14·600	15·400	9·800	9	29·659	17·350	17·275	10·000
36	19·995	16·625	13·875	12·925	38	30·326	15·800	16·100	12·150
3	20·162	15·975	15·875	8·875	2	31·159	15·475	17·750	11·200
5	20·828	14·400	16·100	9·350	23	31·159	17·850	18·325	13·325
20	21·161	16·650	13·950	10·325	30	31·159	15·225	15·225	9·400
6	21·828	15·500	14·175	9·050	4	32·159	13·900	16·000	9·475
25	21·995	16·350	13·950	8·225	27	32·159	16·950	15·125	11·875
13	23·494	16·075	14·625	9·450	1	33·583	13·675	14·925	11·750
12	23·994	16·675	15·575	12·175	7	33·825	14·050	14·250	10·075
16	24·161	12·800	12·175	6·450	35	34·325	13·750	12·225	9·050
29	26·327	15·525	15·200	12·900	31	34·991	15·450	16·150	11·175
24	26·493	15·900	18·425	9·850	8	35·325	14·300	15·975	10·625
40	26·493	14·625	15·300	9·075	33	35·658	19·800	19·375	12·100
10	26·660	16·950	14·225	11·125	28	42·156	13·625	14·825	11·150
21	26·660	15·075	15·950	10·075	26	48·488	16·075	10·125	11·900

Nos. 1–20, secondary modern school subjects. 21–40, E.S.N. school subjects.

Inspection of the size estimates of subjects in the two different school groups shows a considerable difference. Table III gives the details and statistical significance of these differences. While the illusion error of the E.S.N. subjects is substantially larger than the error of the non-E.S.N. children, the difference falls short of statistical significance when performing an analysis of variance (viz. Table VIII) or a Mann-Whitney U-test. A 1-tailed t test gives a significant difference at the 5 per cent. level, but the importance of this significant result is somewhat questionable in view of the analysis of variance result. The between-group differences in size estimation for two of the three standard objects are, however, unequivocally significant. (Generally, as a result of the robustness of non-parametric techniques, t tests of significance are more likely to give significant results as well as a higher level of significance.)

The difference in susceptibility to the Müller-Lyer illusion for groups differing in intelligence suggested the presence of a regular relationship between increasing intelligence and reduction in illusion error. Table IV shows the per cent. illusion error for different intelligence categories of the two major groups. While there is a slight trend for lower error to be associated with higher intelligence, the total

TABLE III

Experimental variables	*Diamond*	*Rod*	*Playing card*	*Per cent. error Müller-Lyer*
Veridical size (in cm.)	15·000	15·000	8·900	—
Veridical size + 5 per cent.	15·750	15·750	9·345	—
Veridical size — 5 per cent.	14·250	14·250	8·455	—
MEANS AND DISPERSION VALUES—				
Non-E.S.N. Group:				
Mean	15·059	15·108	9·831	25·423
s.e.	0·35	0·33	0·30	1·35
S.D.	1·51	1·43	1·32	5·90
E.S.N. Group:				
Mean	16·175	15·270	11·193	29·443
s.e.	0·36	0·51	0·46	1·73
S.D.	1·55	2·22	2·01	7·54
BETWEEN GROUP DIFFERENCE—				
U	133·5	190·5	121	142·5
P (1-tail)	<0·05**	N.S.	<0·025	N.S.*

When using 1-tailed *t* test: * significant at 5 per cent. level; ** significant at $2\frac{1}{2}$ per cent. level.

picture is perhaps better illustrated by a non-significant correlation of —0·192 between the two variables. The difference between non-E.S.N. and E.S.N. samples of equal intelligence is interesting. While the number of subjects involved is small, it would tend to support any conclusion assigning a minor role to general intelligence in this particular task.

TABLE IV

PER CENT. ILLUSION ERROR AND I.Q.

I.Q. range		*61–70*	*71–80*	*81–90*	*91–100*	*101–110*	*111–120*
Per cent. illusion error	Groups						
	Non-E.S.N.	—	—	28·292 (n = 7) *	22·732 (n = 7)	26·027 (n = 5)	21·161 (n = 1)
	E.S.N.	28·077 (n = 6)	28·076 (n = 12) **	41·739 (n = 2)	—	—	—

* p = 0·028, ** p = 0·025 (both 1-tail and Mann-Whitney U tests) for differences between adjacent populations.

The correlations between per cent. illusion error on the one hand, and size estimates and per cent. error of estimates irrespective of direction on the other hand, are shown in Table V. They are uniformly non-significant. Some degree of correlation must be expected, of course, since the fundamental aim of judging the dimensions of objects is the same for both tasks.

Considering size estimates here operationally defined as indicating over-constancy and under-constancy (±5 per cent. veridical standard size), 22 mean estimates fell

TABLE V

	Estimates			Per cent. error estimates		
	Diamond	Rod	Playing card	Diamond	Rod	Playing card
Per cent. Müller-Lyer error	−0·155	−0·015	0·251	−0·058	0·109	0·189

For N = 40, $r = 0{\cdot}313$ for $p = 0{\cdot}05$.

into the former category (8 E.S.N., 14 non-E.S.N.) and 64 mean estimates into the latter (35 E.S.N., 29 non-E.S.N.). For subjects respectively giving over- and under-constancy estimates in response to the three different standards, the per cent. illusion error was calculated. The result is given in Table VI.

TABLE VI

PER CENT. ILLUSION ERROR IN RELATION TO OVER-CONSTANCY AND UNDER-CONSTANCY

Constancy standard		Diamond		Rod		Playing card	
Type of constancy estimate		Over	Under	Over	Under	Over	Under
Per cent. illusion error	Non-E.S.N.	28·711 (n = 5)	22·899 (n = 7)	25·993 (n = 7)	25·535 (n = 8)	21·994 (n = 2)	26·083 (n = 14)
	E.S.N.	38·241 (n = 2)	28·132 (n = 12)	30·366 (n = 5)	28·850 (n = 7)	21·995 (n = 1)	29·961 (n = 16)

None of the differences in illusion error between these over-constant and under-constant sub-groups is statistically significant in the direction confirming the hypotheses, though the differences, particularly in relation to the diamond standard, are in the direction predicted by the constancy scaling hypothesis. A t test was not performed because of the small number in the over-constancy category. It is worth noting, however, that for the standard of known physical dimension it was subjects giving under-constant size estimates who produced the greater illusion error. Combining non-E.S.N. and E.S.N. groups for the calculation, this difference failed narrowly to reach statistical significance at the 5 per cent. level of confidence employing a 1-tailed U test.

Two significant differences are present. For subjects producing over-constancy for the diamond standard, the E.S.N. sample had a significantly larger illusion error than the non-E.S.N. group, and for the E.S.N. sample only, subjects producing under-constancy for the playing card standard had a significantly larger illusion error than the single subject with an over-constant size estimate. Both results were obtained from a U test.

The examples previously given by Gregory to demonstrate the likelihood of the operation of perspective and constancy scaling factors in the perception of the illusion error, dictated the need to consider the co-variation of susceptibility to the illusion and orientation of the experimental figure in addition to the anticipated constant error due to direction of adjustment.

Table VII shows the per cent. error of adjustment for the two groups, for the four positions of the variable in relation to the standard, and for the two directions of adjustment, with an indication of the significance of between-group differences.

An analysis of variance was performed to calculate the effects of directions of adjustment, orientations and interactions.

TABLE VII

Per Cent. Illusion Error, Directions of Adjustment and Orientations of Variable

		Direction		Orientation			
		Reducing	*Increasing*	*"East"*	*"North"*	*"West"*	*"South"*
Groups	Non-E.S.N.	22·610	28·595	22·870	28·728	26·467	24·339
	E.S.N.	28·063	30·723	29·393	29·659	28·994	29·393
Between-group differences * P		<0·01	N.S.	<0·02	N.S.	N.S.	N.S.

* Mann-Whitney 1-tailed U test.

TABLE VIII

Analysis of Variance of Per Cent. Illusion Error

	Sum sq.	*d.f.*	*Mean sq.*	*F*	*P*
Between subjects—					
Groups (G)	653	1	653	3·14	>0·05
Subjects within groups ..	7809	38	206	—	—
Within subjects—					
Directions (D)	835	1	835	37·95	<0·001
D × G	126	1	126	5·73	<0·02
D × subjects within groups	829	38	22	—	—
Orientations (O)	239	3	80	3·08	<0·05
O × G	216	3	72	2·77	<0·05
O × subjects within groups	2909	114	26	—	—
D × O	102	3	34	2·62	>0·05
D × O × G	36	3	12	0·92	N.S.
D × O × subjects within groups	1390	114	13	—	—
Total	15,244	319	—	—	—

Table VIII shows that the effect of directions of adjustment, and orientations of the variable was significant, as were the interactions between directions and groups and between orientation and groups. Of considerable interest is the comparative invariance of the errors for the E.S.N. group, which while it indicates consistency, at the same time shows a relative lack of responsiveness to considerable changes in the perceptual field. In view of this invariance, the difference between different orientations of the variable was considered only in relation to the non-E.S.N. group. As suggested by Table VII, only the difference between the "East" and "North" orientations was significant for both directions of adjustment, and in both cases at better than the 2 per cent. level (Mann-Whitney 2-tail U test). Greater illusion errors appeared when the experimental figure was orientated along the "North-South" axis, but not significantly more. With the figure in the position yielding for the non-E.S.N. group the greatest illusion error ("North"), the correlations

between illusion error and constancy estimates were once again not significant (+0·097, +0·027, +0·286, N = 20, respectively for the three standards).

Discussion

On the basis of these findings, the three hypotheses remain sufficiently unsupported to demand their rejection. No significant correlations were obtained between constancy and degree of illusion, and in relation to the second and third hypotheses partially opposite results were obtained. The present study therefore fails to produce adequate evidence of a common or similar process affecting susceptibility to the Müller-Lyer illusion as well as judgement of size over distance. Instead it has brought out other parameters which ought to be further considered in any discussion of the illusions. Firstly, there is the evidence concerning considerable individual differences in illusion error as well as in constancy judgements. Secondly, there is the evidence that the degree of illusion error is smaller in a group of subjects with high average or greater I.Q.'s. (Evidence reported elsewhere makes it clear that this is not due simply to greater application to task or consistency of the more intelligent subjects.)

A third feature brought out in the present study is that any relationship between size constancy and illusion error may be affected by the nature of the object on which constancy estimates are made. When the standard was the playing card, subjects giving under-constant size estimates produced a near-significantly larger illusion error than subjects with over-constant estimates, while standards of indeterminate dimensions produced a (non-significant) relationship between over-constancy and larger illusion error as predicted by the constancy scaling hypothesis. On the basis of this hypothesis, however, it might be predicted that subjects with under-constancy for objects of known use and dimension should show similarly faulty constancy scaling in relation to the illusion, thus producing smaller illusion errors, but the reverse was found. The constancy explanation is further subject to the deduction that knowledge of perspective and the size-distance relationship predisposes towards greater illusional error. In experiments linked to the one presently reported and which cannot be amplified here, such a deduction is not supported. It was found (Hamilton, 1965) that subjects possessing only primitive or no real knowledge of perspective effects and requirements have significantly greater illusion errors than subjects with adequate perspective knowledge.

Gregory (1965) does not make it clear on how many subjects his elegant evidence linking fin angle, depth estimation and illusion is based or what his instructions were. Since he carried out no separate experiment, it is also not possible to say to what extent over- and under-constancy perceivers were represented in his sample. In addition, on the basis of his own conceptualizations, it might be necessary to control for the precise role of perceptual inference by determining subjects' differential degree of ability or disability to impose perceptual structure upon an ambiguous figure consisting of one (? vertical) line and two angles attached to its extremities, and the effect of any structuring or its absence on the illusion error.

The present study has not been able to make a positive contribution to an understanding of the phenomena of the Müller-Lyer figure which thus continue to defy adequate explanation. The universality of the illusion suggests that it has important adaptive properties, and among the adaptive mechanisms, perceptual and—if the term may be used—neurological inference occupies a high rank. It is possible that this universality is no more than an expression of a general perceptual tendency to apprehend a stimulus in its contextual stimulus field rather than in isolation, a response that has been acquired because it is most frequently appropriate

and reinforcing. This function of visual perception (in advance of a behavioural response) is clearly evident not only in the development of size constancy but also in motivated behaviour and selective perception, and seems to exemplify a "synthesizing" process. In individuals with a lower frequency for this process and its associated responses or with the capacity for its control in appropriate situations, the reality-defying characteristics of the illusions may be partially overcome. The findings of Thurstone and of Gardner, indicating that subjects with greater ability in abstracting limited features of the Gottschald figures have smaller illusion errors, are, therefore, obviously relevant. It is possibly because of deficiencies in this capacity that the present E.S.N. group had larger illusion errors. The difference in illusion error for subjects respectively over- and under-constant in estimating the size of a known object does not seem to be explicable in terms of the evidence at present available.

Dr. D. H. Taylor kindly helped with the computer programmes for this and the associated studies. Special thanks are due to Mr. W. A. Goodworth and Mr. B. H. Tuggey, Headmasters of the schools supplying the subjects, for making research facilities available.

References

Crookes, R. G. (1957). Size constancy and literalness in the Rorschach test. *Brit. J. med. Psychol.*, **30**, 99–106.

Gardner, R. W. (1961). Cognitive controls of attention deployment as determinants of visual illusions. *J. abnorm. soc. Psychol.*, **62**, 120–7.

Gregory, R. L. (1963). Distortion of visual space as inappropriate constancy scaling. *Nature*, **199**, 678–80.

Gregory, R. L. (1964*a*). Illusory perception as a constancy phenomenon. *Nature*, **204**, 302–3.

Gregory, R. L. (1964*b*) Human perception. In *Experimental Psychology: Brit. Med. Bull.*, **20**, 21–6.

Gregory, R. L. (1965). Depth in seeing. *Nature*, **207**, 16–9.

Hamilton, V. (1963). Size constancy and cue responsiveness in psychosis. *Brit. J. Psychol.*, **54**, 25–39.

Hamilton, V. (1965) Size constancy and the Müller-Lyer illusion in stages of cognitive development. Paper to the Experimental Psychology Society Meeting in Reading.

Joynson, R. B. (1958*a*). An experimental synthesis of the associationist and Gestalt accounts of the perception of size. Part I. *Quart. J. exp. Psychol*, **10**, 65–76.

Joynson, R. B. (1958*b*). An experimental synthesis of the associationist and Gestalt accounts of the perception of size. Part II. *Quart. J. exp. Psychol.*, **10**, 142–54.

Oyama, T. (1960). Japanese studies on the so-called geometrical-optical illusions. *Psychologia*, **3**, 7–20.

Perry, P. (1961). Size constancy in normals and schizophrenics. In Ittelson, W. H. and Kutash, S. B., (Ed.) *Perceptual Changes in Psychopathology* New Brunswick, N.J.

Raush, H. L. (1952). Perceptual constancy in schizophrenia. *J. Personal.*, **21**, 176–87.

Rump, E. E. (1961). The relationship between perceived size and perceived distance. *Brit. J. Psychol.*, **52**, 111–24.

Tausch, R. (1954). Optische Tauschungen als artifizielle Effekte der Gestaltungsprozesse von Grössen-und Formenkonstanz in der natürlichen Raumwahrnehmung. *Psychol. Forsch.*, **24**, 299–348.

Tausch, R. (1955). Nichtbewusste (sog. unbewusste) Vorgänge bei der optischen Grössenwahrnehmung von Gegenständen. *Psychol. Forsch.*, **25**, 28–64.

Thouless, R. H. (1931*a*). Phenomenal regression to the real object. I. *Brit. J. Psychol.*, **21**, 339–59.

Thouless, R. H. (1931*b*). Phenomenal regression to the real object. II. *Brit. J. Psychol.*, **22**, 1–30.

Thouless, R. H. (1932). Individual differences in phenomenal regression. *Brit. J. Psychol.*, **22**, 216–41.

Thurstone, L. L. (1944). *A Factorial Study of Perception.* Chicago.

Vernon, M. D. (1952). *A Further Study of Visual Perception.* Cambridge: University Press.

Manuscript received 21st August, 1965.

COMMENT ON DR. VERNON HAMILTON'S PAPER

BY

R. L. GREGORY

From the Department of Experimental Psychology, University of Cambridge

Dr Hamilton has, unfortunately, omitted the vital feature of my theory of visual illusions; namely, that constancy is not set only according to apparent distance but that it is also set directly by depth features, especially perspective. In illusion figures we find perspective depth features on a flat plane: if constancy were set by these features it must be inappropriate, because there is no actual depth in the figure, and it would produce size distortions. This is the basis of the theory, which I will now describe in sufficient detail to assess the importance of Hamilton's low correlations between normal constancy and the illusions.

The notion that the illusion figures can be regarded as perspective projections of corners and parallel lines goes back to the "Perspective Theory" (Woodworth, 1938). This theory, however, failed to show why illusory *expansion* should occur with perspective *distance*. Several writers have suggested that constancy may be involved, and this is attractive and plausible for several reasons: (1) Constancy size scaling certainly occurs, and so we do not need to postulate some unknown process for the distortions if it could be responsible. (2) Constancy is normally related to distance, and so it must somehow be set by distance information. This must include some monocular depth cues, since constancy functions with a single eye, and could include perspective. (3) The direction of distortions goes the right way if we suppose that perspective features set constancy, for constancy requires *expansion* with distance to compensate the geometrical shrinkage of retinal images with increasing distance. It is, however, entirely clear that if the distortion illusions are to be attributed to constancy mechanisms we must not think of constancy in the traditional way as simply tied to apparent distance—because the illusion figures appear both distorted and flat at the same time.

Perspective figures present the eye with a special and peculiar problem. They present perspective depth information on a flat plane. As objects in their own right they lie in two dimensions, and yet they represent objects lying in a quite different three-dimensional space. The amazing thing is that we manage to make any sense out of such a paradoxical situation as two and three dimensions combined. It is not surprising that perspective drawing is a late invention, and requires special perceptual learning.

In this queer case of perspective figures, constancy must either: (1) follow the depth indicated by the perspective features of the picture, or (2) it must follow the depth of the plane on which the picture lies. Size scaling cannot possibly be appropriate to both sets of distances at once! The fact that we get systematic expansion with the figure's perspective (the distortion illusions) indicates that size scaling is directly affected by the picture's perspective, even when the picture is seen lying on its background plane. Such perspective-induced size scaling *must* be inappropriate, for the depth features of the picture do not correspond to its true depth, since it is actually flat.

Direct setting by perspective features ("Primary Scaling") is not, however, the whole story of constancy: constancy can also be set according to apparent distance. This is not, however, shown clearly by the classical Emmert's law after-image situation, for here the retinal image changes with change in the background distance. It is thus possible that changing depth features of the background set the size-scale of the after-image directly. But, in fact, constancy scaling can occur with no change in retinal stimulation: this is shown by the behaviour of luminous depth-ambiguous figures. These change in shape with depth reversal; the apparently more *distant* face of, for example, a luminous Necker cube, always appearing the *larger*. But since this size scaling occurs with no change in the retinal image (the figure may even be optically stabilised on the retina) apparent size cannot here be set directly by perspective or other retinal depth cues. It must be set by the central, and changing, interpretation or "hypothesis" of how the parts lie in depth, not directly by depth features of the figure, for these remain unchanged at the retina.

Although luminous depth-ambiguous figures do change in size with seen depth, ambiguous figures do not change in size *when presented on textured backgrounds*. Perceptual reversal in the paradoxical depth of pictures on visible backgrounds does not

affect constancy scaling. Now illusion figures presented on a background can reverse in paradoxical depth, but the illusion remains unchanged by the reversals. When shown as *luminous* figures with no background, however, they *do* change in size, following Emmert's law. The presence or absence of a background is thus all-important, though generally ignored.

The general picture is that apparent size is determined by two scaling processes. The "Primary Scaling" acts directly from local depth cues, and is unaffected by interpretation or apparent depth. "Secondary Scaling" works from apparent distance, depending entirely on non-paradoxical seen distance. In other words: Primary scaling works directly from available depth *data*; secondary scaling from the prevailing perceptual *answer*—seen depth. Either or both of these may be inappropriate to the true depth, and either will produce distortions when inappropriate. Normally they both serve to compensate for the shrinkage of retinal images with distance, but in perspective drawings the situation is logically paradoxical, and they *must* give inappropriate scaling. The distortion illusions are not arbitrary, or contingent upon how brains are made (a seeing machine would be similarly affected). In this queer situation of flat depth-perspective pictures, we experience the illusions dramatically because the constancy evoked by the perspective on a flat plane is highly inappropriate. Illusions are not however limited to the case of perspective figures: errors in either primary or secondary scaling must produce systematic size-distortions and either can be wrongly set in the normal world, though fortunately this is rare.

Now the question is: should we expect a high correlation in individual differences between these illusion distortions and normal size constancy? As should now be clear, Hamilton is wrong in saying that I infer that "the perceptual behaviour in relation to the illusion is . . . similar . . . to that operating in the estimation of size over distance, and that two parts of the Müller-Lyer figure will in fact be perceived as, say, the corners of a house or a room. . . ." What I do say is that this is only the case when the background is invisible. When it is visible they appear flat. (They do not *have* to be seen in depth even when it is invisible, for the corner is no more than the *preferred* depth interpretation.) Hamilton ignores my primary scaling, but this is essential to the theory for the whole point is that size scaling is *not* tied to seen depth in perspective (illusion) figures with textured backgrounds. It *is* tied to depth cues, even though inappropriate because the figures are actually flat—hence the illusion.

In normal conditions, it is mainly secondary scaling which sets apparent size. Primary constancy plays but a relatively minor role as we see from experiments on depth-ambiguous figures and from the limited distortion (about 30 per cent. maximum) found in the illusion figures. This represents but a small proportion of constancy scaling where depth is non-paradoxical, and secondary scaling is effective. But in the illusion figures it is primary scaling which is responsible for the distortions, for the figures are seen as flat and so secondary scaling cannot function. Thus: (*a*) secondary scaling is mainly responsible for constancy for normal objects (and especially with stereoscopic vision, which is, however, not effective for flat figures); (*b*) primary scaling is entirely responsible for the distortion illusions presented on textured backgrounds. Since the effective processes are different, we should not expect any high correlation between constancy for normal objects and distortions which occur when three-dimensional information is presented on a flat plane—to confound the eye with a paradox.

Reference

Woodworth, R. S. (1938). *Experimental Psychology.* New York: Holt.

Manuscript received 23rd October, 1965.

SHORTER ARTICLES AND NOTES

THE EFFECT OF OVERTRAINING AND REPEATED EXTINCTIONS ON SPEED OF EXTINCTION IN ELECTRICAL SELF-STIMULATION

BY

L. J. HERBERG and P. J. WATKINS

From the Institute of Neurology, Queen Square, London, W.C.1

Fourteen rats with hypothalamic electrodes needed at least 3,000 reinforcements in self-stimulation before extinction scores reached their peak. In a series of up to 30 further extinctions and reconditionings, involving a total of 10,500 reinforcements, extinction scores fell to the extreme low values typical of self-stimulation. The extent of this fall was shown to depend on the number of extinctions administered, but not on the number of reinforcements, indicating that a process of conditioned inhibition may be partially responsible for rapid extinction after self-stimulation.

Introduction

In urging that electrical self-stimulation should not be ascribed to the effects of normal reinforcement, Kimble (1961, p. 263) states that "extinction (after self-stimulation) is much more rapid than . . . in the usual bar-pressing situation"—an impression which is supported by the sharply-breaking extinction curves seen in typical cumulative records (e.g. Olds, 1955). However, it is possible that this rapid extinction merely reflects features of the reinforcement schedule rather than any peculiarity of the reinforcement itself, and the present report extends an earlier study supporting this view (Herberg, 1963). This earlier study showed that extremely rapid extinction could be produced as a training effect by administering a long series of unvarying bouts of self-stimulation, and it was concluded that the rapid extinction was caused either by practice in extinction leading to progressively lower scores (e.g. Perkins and Cacciopo, 1950; Bullock and Smith, 1953), or by an overtraining effect arising from the several thousand reinforcements administered during the course of training. Overtraining might be expected to lead to this result because it has been shown to lessen resistance to extinction in runways after partial (Capaldi, 1958) and continuous (North and Stimmel, 1960) reinforcement, and some workers (King, Wood and Butcher, 1961) but not others (Margulies, 1961; D'Amato, Schiff and Jagoda, 1962) have reported the same effect under free operant bar-pressing conditions similar to those imposed in self-stimulation. But in seeming contradiction to the extension of these findings to self-stimulation was Herberg's observation that some subjects gave low extinction scores even on the first day of self-stimulation; however, these low scores may have reflected an early overtraining effect produced by the several hundred reinforcements necessarily administered in the course of "shaping." This possibility was tested in the first of the following experiments by giving the subjects a deliberately prolonged shaping session to see if the extra reinforcements would lead to even lower scores. The second experiment compared the effects on extinction, of over-training, and of "training to extinguish."

Experiment 1

Method

Fourteen rats that showed regular hypothalamic self-stimulation were shaped in the usual way except that after they had learnt to operate the pedal, they were allowed to make 1,500 additional responses before the stimulating current was switched off. The number of unrewarded responses made in the next 2 min. provided an initial extinction score, which was compared to the initial scores of a control group of 15 consecutive subjects with hypothalamic electrodes, used in previous experiments. For the control

group the shaping sessions had been comparatively brief, i.e. never including more than 500 reinforcements (range 111–410; mean 207). Self-stimulation rates for all rats were never less than 40 responses per minute.

No subject was included if it broke its leads while responding in this or the following experiment, but it was sometimes necessary to give one or two priming shocks to those that had stopped responding because of seizures. Since interruptions by seizures tend to occur only at the beginning of sessions (Herberg and Watkins, 1966), they probably had little effect on extinction scores. Very rarely, some rats were found to be still responding, in a sporadic way, within 30 sec. of the end of the 2-min. extinction period, and when this occurred the extinction periods were prolonged until 30 sec. had passed without a response. But only those responses made during the first 2 min. were counted.

Results and discussion

The mean extinction score of the 14 experimental rats was 18·9 (range 7–48), and that of the control sample was 6·9 (range 1–27). The difference is highly significant (Mann-Whitney $U = 21 \cdot 5$; $p < 0 \cdot 002$), but its direction shows that the extra reinforcements given to the experimental sample tended to increase resistance to extinction, not reduce it. Thus the low scores previously found after shaping were probably caused by undertraining rather than overtraining, with extinction being imposed before acquisition was complete. This conclusion was confirmed in the next experiment, in which the experimental rats were given differing amounts of further reinforcement during a series of repeated extinctions and reconditionings.

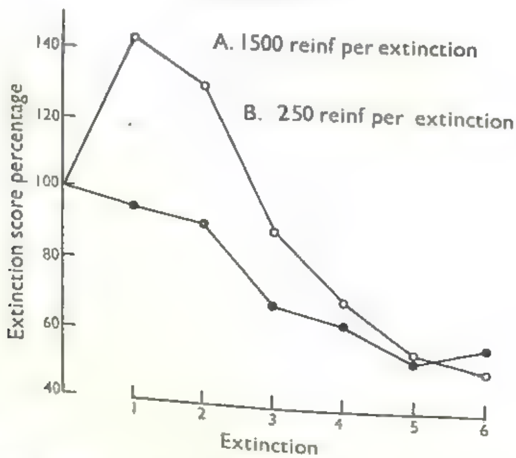
Experiment 2

Method

The 14 rats used in Experiment 1 were randomly assigned to two equal groups. Group A was allowed to make 1,500 reinforced responses in a continuous session on each of six more days, and a 2-min. extinction score was recorded at the end of each session. Group B also made 1,500 responses daily, but sessions were divided into six bouts of 250 responses, each bout being followed by a 2-min. extinction period. The extinction scores of Group B were recorded after each of the six bouts given on day 1, and after every sixth bout on the five following days.

To permit the trends of different animals to be compared, each animal's extinction score was expressed as a percentage of its initial extinction score recorded in Experiment 1. The mean initial scores of the two groups were A = 20·6 and B = 17·1, the difference being non-significant (Mann-Whitney $U = 25 \cdot 5$; $p = 0 \cdot 5$).

FIGURE 1



Consecutive extinction scores of two groups of subjects given a differing number of reinforcements. Each point is the mean percentage score of seven rats.

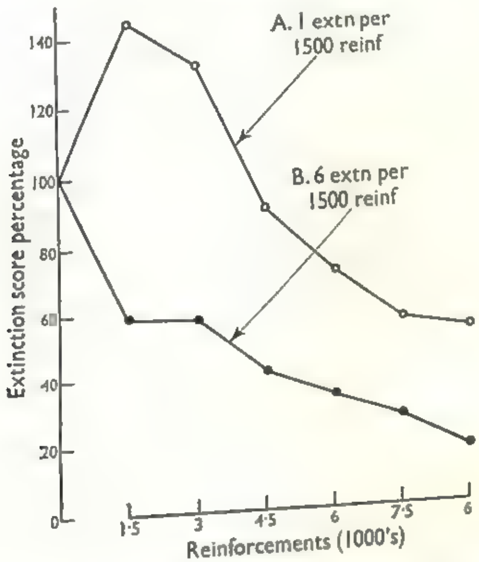
Results and discussion

Figure 1 compares the effects on extinction of different amounts of reinforcement. The scores of both groups showed an overall decline during the course of successive extinctions, but except during the first two extinctions there was no difference between the scores of rats receiving 1,500 reinforcements, and those receiving only 250 reinforcements per extinction. Even after a total of 10,500 reinforcements (including those in Experiment 1) had been given to Group A, there was no evidence that scores of this group had been affected by overtraining. This failure to demonstrate an overtraining effect in the Skinner box agrees with the results reported by Margulies (1961) and D'Amato, Schiff and Jagoda (1962) rather than with the findings of King, Wood and Butcher (1961).

The elevation of the scores of Group A on the first two days was shown to be significant in a Friedman two-way analysis of variance ($\chi_r^2 = 8{\cdot}3$; $p < 0{\cdot}05$). This finding, together with the results of Experiment 1, shows that at least 3,000 reinforced responses may be necessary before acquisition in self-stimulation can be said to be complete.

Figure 2 shows the effects of different amounts of extinction training in subjects receiving the same number of reinforcements. Each sixth extinction score of Group B

FIGURE 2



Extinction scores of two groups of subjects given the same number of reinforcements but a differing number of extinctions. For Group B, every sixth extinction is plotted.

was significantly lower than each successive score of Group A (Mann-Whitney $U < 8$; $p < 0{\cdot}05$). Since the only difference in the treatment of the two groups was that Group B underwent six times as many extinctions, it follows that the common practice of administering self-stimulation in a series of repeated bouts must be responsible, at least in part, for the abrupt extinctions that usually follow. The phenomenon can be readily explained as an instance of conditioned inhibition, and the prediction follows that disinhibition by any strong and unexpected extraneous stimulus during extinction would lead to renewed bar-pressing (Hilgard and Marquis, 1960, p. 127). Such a result has indeed been reported by Deutsch and Howarth (1962), but the present finding seems to offer a more parsimonious account of the result than the supposed evocation, by fear, of a drive to self-stimulate.

This work was supported by grants from the Medical Research Council and the Institute of Neurology's Rockefeller Research Fund.

References

Bullock, D. H., and Smith, W. C. (1953). An effect of repeated conditioning-extinction upon operant strength. *J. exp. Psychol.*, **46**, 349–52.

Capaldi, E. J. (1958). The effect of different amounts of training on the resistance to extinction of different patterns of partially reinforced responses. *J. comp. physiol. Psychol.*, **51**, 367–71.

D'Amato, M. R., Schiff, D., and Jagoda, H. (1962). Resistance to extinction after varying amounts of discriminative or non-discriminative instrumental training. *J. exp. Psychol.*, **64**, 526–32.

Deutsch, J. A., and Howarth, C. I. (1962). Evocation by fear of a habit learned for electrical stimulation of the brain. *Science*, **136**, 1057–8.

Herberg, L. J. (1963). Determinants of extinction in electrical self-stimulation. *J. comp. physiol. Psychol.*, **56**, 686–90.

Herberg, L. J., and Watkins, P. J. (1966). Epileptiform seizures induced by hypothalamic stimulation in the rat: resistance to fits following fits. In press.

Hilgard, E. R., and Marquis, D. G. (1940). *Conditioning and Learning.* New York: Appleton Century-Crofts.

Kimble, G. A. (Ed.) (1961). *Hilgard and Marquis' Conditioning and Learning.* London: Methuen.

King, R. A., Wood, P., and Butcher, J. (1961). Decreased resistance to extinction as a function of reinforcement. *Amer. Psychologist*, **16**, 468.

Margulies, S. (1961). Response duration in operant level, regular reinforcement, and extinction. *J. exp. Anal. Beh.*, **4**, 317–21.

North, A. J., and Stimmel, D. T. (1960). Extinction of an instrumental response following a large number of reinforcements. *Psychol. Rep.*, **6**, 227–34.

Olds, J. (1955). Physiological mechanisms of reward. In Jones, M. R. (Ed.). *Nebraska Symposium on Motivation.* Pp. 73–139. Lincoln: Nebraska University Press.

Perkins, C. C., and Cacciopo, A. J. (1950). The effect of intermittent reinforcement on the change in extinction rate following successive reconditionings. *J. exp. Psychol.*, **40**, 794–801.

Manuscript received 29th June, 1965.

THE ROLE OF STIMULATION IN THE DELAY OF ONSET OF CRYING IN THE NEWBORN INFANT

BY

T. GORDON and B. M. FOSS

From the Department of Psychology, Birkbeck College, and the Institute of Education, University of London

Crying is said to result from stimulus need. An experiment was done in which infants were stimulated during quiet periods to see if this would postpone the onset of crying. The results were positive. An alternative explanation in terms of internal temperature control is considered.

Introduction

Crying in the human infant has been treated either (1) as a reaction to an unpleasant external stimulus, or (2) as a response to an internal need. Point (1) is illustrated by Skinner's observation (1959, p. 420) that he was able to reduce the amount of crying of his own child by keeping it in a compartment in which the temperature was held constant and the child could move about without the constraints of clothing. Point (2) is illustrated by the terms "stimulus need" and "stimulus hunger" which have been used by investigators who are concerned with the kind of stimulation required for normal development in both humans and animals (Thompson, 1955, p. 133). Presumably the infant requires varied stimulation, the kind and quantity of which changes as the child grows. Crying can be regarded as an index of the absence of such stimulation.

Crying occurs fairly regularly in the human infant before feeding. It also occurs, though less frequently, when there is no obvious indication of discomfort: the child need not be windy, wet or in discernible pain. In the newborn infant, one of the first efforts which the mother makes to stop the infant from crying is to rock it. Sometimes this is successful and the infant drops off to sleep again. By analogy with hunger, the need involved is reduced. If it is not other efforts will be made to find out why the baby is crying and how to stop it.

Since rocking seems an effective means of stopping the baby from crying when it is not hungry or in pain, it would appear to answer to some of the stimulus needs of the young infant. If it can be effective in stopping the infant from crying, rocking the child before it cries should prevent or delay the infant's crying effectively. That is to say, if crying can result from stimulus hunger then appropriate stimulation should put off the onset of crying. The following experiment was done to test this hypothesis.

Experimental Method

The experiment was carried out in the nursery of one of the wards of an obstetrics hospital. All the children here were full term babies. They were placed in the nursery for about 1½ hours after their noonday feed, the only time during the day when they were together and away from the bedsides of their mothers.

Each day the number of babies in the nursery who were not crying was counted and one baby was selected for rocking; the remaining babies who were not crying formed the control group. The experimenter stood at the foot of the cot and rocked the baby in it for half an hour. After this the experimenter remained in the nursery for half an hour. The number of control babies who cried during the hour the experimenter was present was noted.

The cots were numbered by the experimenter and each day the number of the cot selected for rocking was moved up by one. If the baby whose turn it was to be rocked was already crying, the number was moved up by one or to the next baby who was quiet.

The experiment was carried out for 18 days between 25th May and 24th August, 1964. The maximum number of babies in the nursery on any one of these days was 11; the minimum number was seven. The maximum number of babies who could be used for the experiment, that is, the number of quiet babies, was 10; the minimum number of quiet babies was six. The ages of the babies ranged from a few hours to 10 days.

Results

The results are listed in Table I.

TABLE I

Day	*Control babies*	*Control babies who cried*	*Experimental babies*	*Experimental babies who cried*	*Proportion control babies not crying*	*Proportion experimental babies not crying*	*Difference in proportion*
1	8	5	1	0	0·37	1·00	0·63
2	6	4	1	0	0·33	1·00	0·67
3	5	4	1	0	0·20	1·00	0·80
4	6	5	1	1	0·16	0	−0·16*
5	5	1	1	0	0·80	1·00	0·20
6	9	5	1	0	0·44	1·00	0·56
7	8	3	1	0	0·62	1·00	0·38
8	8	4	1	0	0·50	1·00	0·50
9	5	2	1	0	0·60	1·00	0·40
10	9	1	1	1	0·88	0	−0·88
11	6	1	1	0	0·83	1·00	0·17
12	9	1	1	0	0·88	1·00	0·12
13	8	3	1	0	0·62	1·00	0·38
14	5	1	1	0	0·80	1·00	0·20
15	6	2	1	0	0·66	1·00	0·34
16	8	1	1	0	0·87	1·00	0·13
17	6	2	1	1	0·66	0	−0·66
18	8	3	1	0	0·62	1·00	0·38

* On this day the outdoor temperature had dropped from 72° F. to 52° F. The heating in the nursery was not turned on until very close to the end of the observation period.

The null hypothesis, that the experimental babies did not differ from the control babies, would be substantiated if the differences in proportion were not significantly different from zero. A t-test gave a probability of approximately 0·02 (1-tailed). A correction for skewness increases the probability to between 0·03 and 0·04.*

Discussion

The results of the experiment give reasonable evidence that rocking answers some need of the newborn infant since it not only stops crying which has begun but effectively delays its onset. Two possible explanations for what this "need" is present themselves. It could be (1) that the newborn infant requires kinaesthetic stimulation or (2) that further, this kinaesthetic stimulation is a means of helping to regulate the internal temperature of the infant.

Before the experiment was begun a period of time was spent observing the babies on the ward and in the nursery. From these observations there seems little doubt that, once crying has begun, lifting and holding the infant is a more reliable way of stopping its crying than is rocking. Had it been practicable to use holding as the experimental variable the incidence of crying in the experimental babies might have been even further reduced. Lifting and holding the infant might be more effective in stopping crying because the positional change is more marked. Equally, or more important, may be the warmth which the infant gets from the contact with the adult's body. Since internal temperature regulation is not stable in the first weeks of life a lowering of the internal temperature may well account for the onset of crying in what otherwise appear to be periods free from discomfort. Unfortunately, in this experiment the temperature in the

* We are indebted to Professor David R. Cox, of Birkbeck College, for suggesting this method of analysis. His paper based on these data will appear in *Biometrika*, Vol. 53, June, 1966.

nursery was not recorded until the eighth day. However, the outdoor temperature was fairly steady in the first three days (70° F. on day 1, 70° F. on day 2 and 72° F. on day 3). On day 4 the temperature had dropped 20 degrees, to 52° F. and the nursery was unheated until close to the end of the experimental period. The high incidence of crying on that day (only one baby remained quiet) lends some support to the explanation that internal temperature affects the onset of crying.

The crying itself is vigorous and involves movement of the whole body of the infant, especially its head and limbs. One might therefore regard it as the infant's means of attempting to restore its internal temperature or of supplying its need for positional change. However, the infant is very limited in its capacity for movement and in its capacity to restore its temperature equilibrium. Crying would therefore appear to occur when the infant is in need of stimulation which it cannot adequately provide for itself. In this sense it is the infant's chief means of communication about its needs. Under normal conditions the mother's response to her crying infant will ensure that whatever alterations are required will be made.

It had been considered possible that the cessation of rocking might be frustrating to the infants, and lead to crying. In the event, any such effect was overridden.

We should like to thank Prof. Bonham-Carter, Matron Billings and Sister Lee of University College Hospital, London, for making available facilities for this study.

References

SKINNER, B. F. (1959). *Cumulative Record*. London: Methuen.

THOMPSON, W. R. (1955). Early environment—its importance for later behaviour. In HOCH, P. H., and ZUBIN, J. (Ed.). *Psychopathology of Childhood*. New York.

Manuscript received 14th August, 1965.

APPARATUS

A TRANSISTORISED VOICE-OPERATED RELAY

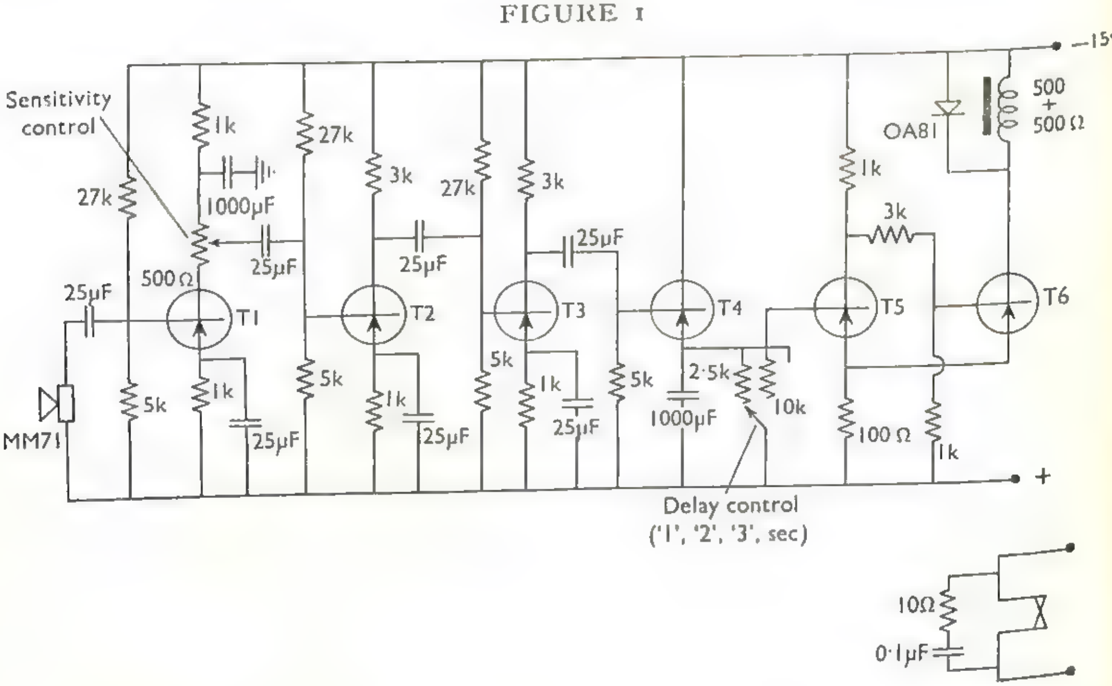
BY

P. G. M. DAWE

From the Institute of Experimental Psychology, Oxford University

There is a frequent need in speech experiments involving the switching of tape recorders, or in measuring voice reaction times, for a voice operated key switch which will close a pair of contacts on receiving a voice signal, switching on a tape recorder or other apparatus, and open them after a given delay, when the speech signal is removed.

FIGURE 1



T1-T6 : OC75 Voice Key Circuit

Although the circuit of such a switch is fairly straightforward (Fig. 1) there are several design points worth noting, or it may be assumed that such a key can be put together without much thought, and unsatisfactory results may be obtained.

Requirements

The main requirements are first, for a simple audio amplifier (T1 to T3) capable of producing sufficient gain to operate a trigger circuit from an inexpensive microphone input (to preserve the quality of the speech is not important) of about 1 millivolt r.m.s. The output from the amplifier must be rectified and be capable of producing several volts of d.c., sufficient to operate the trigger which will open or close the contacts of a high speed relay, of the Siemens type. The amplifier can be a simple three-stage transistor audio amplifier, and the rectification can conveniently be done in an emitter follower circuit (T4). This emitter circuit should contain a large capacitor, sufficient to give a delay in the switching when the audio input is removed. This will prevent the voice key from opening between the words of a sentence, since although the capacitor can be charged quickly, within a millisec., through the forward base-emitter path of the transistor, it will only discharge slowly through the input impedance of the following trigger stage.

The amplifier

The steps in the design of a suitable transistor audio amplifier to provide a gain of several thousand times from 1 mV. r.m.s. to several volts a.c. are as follows. One may assume the input of a normal single grounded emitter stage to be nominally 1,000 ohms (Amos, 1961). The emitter bias resistor may also be fixed at this value since this will only lose a volt or two of the collector supply voltage, for a nominal collector current of a few milliamperes. This resistor will be by-passed by a large capacitor (25 μF.) to prevent loss of gain through local negative feedback. The potential divider for the base bias follows conventional design practice, the bleed current being about ten times the base current and one-fifth of the collector current. These values come out at approximately 27 k-ohm and 5 k-ohm. If a collector load resistance of 3 k-ohm is selected, the effective load resistance, R_c, of the stage will be given by:

$R_c = \frac{3 \times 1}{3 + 1} \times 1{,}000 = 750$ ohms, reckoning the input resistance of the following stage as 1 k-ohm. If I_c is the collector signal current, we have for each stage that:

$$\text{Signal volts output} = I_c R_c$$

$$\text{Signal volts input} = \frac{I_c r_i}{h_{fe}}$$

and hence for an OC75 transistor where h_{fe} is approximately 50, and where $R_c = 750$ ohms, and $r_i = 1$ k-ohm, the stage gain is given as:—

$$\frac{50 \times 750}{1{,}000} = 37{\cdot}5 \text{ times.}$$

Two stages of this amplification, preceded by a pre-amplifier stage with a lower collector load, of say, 500 ohms, to give additional gain of some 10–12 times, provides an audio amplifier with a maximum overall gain of greater than 12,000 times. The 500 ohm load in the pre-amplifier stage can conveniently be made by a potentiometer, and will operate as a sensitivity control for the voice key. The overall gain is then sufficient to operate the trigger circuit from a moving iron type of microphone having a nominal impedance of 1,000 ohms to match the transistor input circuit. The microphone can be laid flat on the table and the amplification is adequate to operate the key from speech or other sounds made within 8 or 10 ft. of the table.

With the microphone fully loaded, the input to the emitter follower rectifier circuit is normally a square wave of some 6 volts amplitude (peak-to-peak) for a 15-volt supply voltage. This waveform charges the 1,000 microfarad emitter load capacitance to a maximum d.c. voltage of 3 volts, or to about 2 volts when driving into the input impedance of the following Schmitt trigger circuit.

The Schmitt trigger circuit

The design of the Schmitt circuit (T5 and T6) starts from the requirement of a collector current change of 12 milliamperes needed to operate the 1 k-ohm relay load. Both collector load resistors can be made the same value, i.e. 1 k-ohm each. If the common emitter load resistor is then made 100 ohms, this will give 1·2 volts emitter bias, for the 12 milliampere collector load current. The first transistor will turn on when its base voltage exceeds this value, i.e. rises towards 2 volts negative on the receipt of an audio input signal. The potential divider values between the first and second transistors must be such as to ensure that the second transistor is switched off, when the first is on. Values of 3 k-ohm and 1 k-ohm are satisfactory. With the relay in the collector circuit of the 2nd transistor it will be opened by the receipt of a signal, closing a pair of contacts. It will then also be open if the voice key is turned off thus ensuring that the external circuit is normally made when the key is not operating.

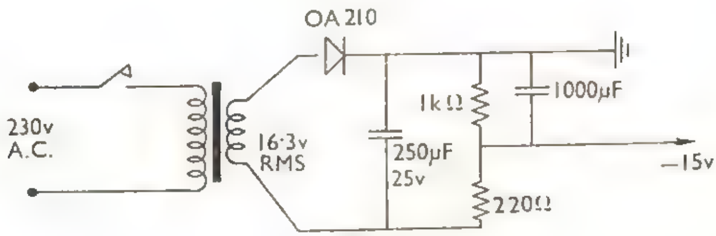
The input impedance of the emitter coupled Schmitt trigger circuit will be a few thousand ohms for a common emitter resistance of 100 ohms and this input impedance can form a time constant of several seconds with a capacitance of the order of 1,000 microfarad, and provide a suitable switching off delay. Several different values of resistance were shunted across the trigger input, via a rotary switch, to make this delay time variable.

Power supply

The power supply for the unit can be either from a battery or from a ½ wave a.c. mains rectifier. If batteries are used an 18 volt supply (2 × DT9 batteries) is recommended,

since although the current drain is about 25 milliamperes, the unit will operate with voltages down to about 10 volts, and so a reasonably long battery life is ensured. A simple a.c. mains source was made from a 16·3 volt r.m.s. transformer, half-wave rectified, through an OA 210, and smoothed with 1,000 microfarad. A potential divider dropped the supply voltage to 15 volts so as not to exceed the collector supply voltage of the OC 75 transistors. It was found advisable to include decoupling (1 k-ohm and 1,000 μF.) in the preamplifier stage to prevent unwanted a.c. signals from getting into the amplifier and operating the key. The circuit of the power supply is given in Figure 2.

FIGURE 2



Power Supply

The Siemens high speed relay (H 85B) contact is suitable for carrying 1 ampere at 50 volts, and a spark quench of 0·1 microfarad and 10 ohms was connected in series across the relay contacts.

Construction

The instrument can conveniently be put together in a small case (10 in. × 6 in. × 6 in.) The transistors and other components can be mounted directly on to a strip of veroboard. This can be mounted on to an aluminium plate secured to the front panel. The use of transistors rather than valves makes for a lighter and more compact unit with a simpler and less bulky power supply, and one which may be assembled within a week or two.

REFERENCE

AMOS, S. W. (1961). *Principles of Transistor Circuits.* London, pp. 56, 89, 90.

Manuscript received 13th March, 1965.

PROCEEDINGS OF THE EXPERIMENTAL PSYCHOLOGY SOCIETY, 1965

29th–31st March, 1965. Extended Meeting at Reading.

1st Session: "Size constancy and the Müller-Lyer illusion in stages of cognitive development," by Dr. V. Hamilton.* "Serial rote-learning: A challenge to S-R orthodoxy," by Professor A. R. Jensen. "Electro-cerebral signs of expectancy, attention and decision," by Dr. W. Grey Walter.*

2nd Session: "Sleep, dreams, drugs and nightmares," by Dr. I. Oswald. "Four-eared listening and two-handed speaking," by Dr. N. Moray.

3rd Session: "Reaction time distributions and classification models," by Mr. D. H. Taylor* and Dr R. Davis. "Effect of fore-period distribution on simple reaction time," by Dr. R. J. Audley, A. R. Pike* and I. B. Stilitz.* "Choice reaction with variable S-R mapping," by Dr. H. Shaffer.*

4th Session: "Contrast discrimination in stabilized vision," by Dr. D. C. West.* "The pulse response of the eye," by Dr. P. R. Boyce.*

7th–9th July, 1965. Extended Meeting at Oxford.

1st Session: "A perceptual maze test: Some parameters determining problem solving," by Mr. G. Jessup* and Dr. A. Elithorn. "Computer simulation of the learning machine STeLLA in a variety of environments," by Mr. J. H. Andreas.* "Forebrain stimulation and feeding behaviour in the Barbary dove," by Dr. D. Vowles.

2nd Session: "Sleep duration," by Dr. W. B. Webb. "Sleep and monotonous stimulation," by Dr. B. Tizard. "Control of heart rate," by Dr. J. Brener* and Mr. D. Hothersall.*

3rd Session: Symposium on Memory convened by Dr. J. Brown.—"The acoustic element in short term memory," by Dr. A. Baddeley. "Application of signal detection theory to memory," by Dr. M. Treisman, Dr. A. Treisman, Mr. G. Geffen* and Mr. T. Walsh.* "The effect of prior recall on multiple response recognition," by Dr. J. Brown and Mr. D. Packham.* "Response duration and response distribution in avoidance schedules with different types of warning signal," by Dr. H. Hurwitz and Mr. P. V. Dillow.*

4th Session: "Attention and probability learning," by Dr. N. J. Mackintosh. "The role of attention in the disinhibition of displacement activities," by Dr. D. J. McFarland.*

5th Session: "On the latency of preparation," by Dr. P. Bertelson. "Problems in the transfer of training," by Dr. M. Hammerton.

5th–7th January, 1966. Annual General meeting at Birkbeck College, London.

1st Session: "An investigation of variables associated with learning impairments following temporal lobectomy in man," by Mr. C. B. Blakemore.* "Artificial learning," by Dr. D. H. Holding. "Judgement time as a test of neural quantum theory," by Mr. T. Shallice.*

2nd Session: "Does the passive mean the same as the active," by Mr. P. N. Johnson-Laird." "Stimulus selection during simple discrimination learning by humans," by Professor G. Bower.*

3rd Session: "Differential effects of R.I. in S.T.M. of acoustic semantic similarity," by Dr. H. C. A. Dale and Dr. M. Gregory.* "The prediction of individual differences during acquisition of a bar pressing situation," by Dr. D. Joyce* and Professor A. Summerfield. "Analysis of cross modal effects and their relationship to language," by Dr. G. Ettlinger.

4th Session: "Individual differences in apparent motion after effects, motion sickness susceptibility, and loudness estimation," by Mr. J. T. Reason.* "The effects or changes in apparent weight upon discrimination and response bias," by Miss H. Ross.*

5th Session: "Linguistic rules and speech perception in schizophrenia," by Mr. D. Gerver.* "Immediate memory attention and ageing," by Dr. F. I. M. Craik.*

* By invitation.

Committee, 1965: President: Dr. L. Weiskrantz.

Editor: Professor O. L. Zangwill.

Ordinary members of the Committee: Dr. A. Baddeley, Dr. R. Davis (Hon. Treasurer), Professor I. M. L. Hunter, Dr. M. Kinsbourne, Mr. D. Legge (Asst. Hon. Secretary), Professor R. L. Reid, Professor J. Symons, Dr. P. H. Venables (Hon. Secretary), Dr. G. K. Wallace.

The following have been elected into membership of the Society:

K. J. Connolly, B.Sc., Department of Psychology, University of Sheffield, Sheffield, 10.
J. A. Gray, Ph.D., Institute of Experimental Psychology, 1 South Parks Road, Oxford.
A. H. Gregory, Ph.D., Department of Psychology, The University, Oxford Road, Manchester, 13.
D. H. Howes, Ph D., Department of Neuropsychology, Boston University School of Medicine, 80 East Concord Street, Boston, 18, Mass.
A. W. Still, Ph.D., Department of Psychology, The University, 7 Kepier Terrace, Gilesgate, Durham.
D. H. Taylor, Ph.D., Department of Psychology, The University, Reading.
P. Whittle, Ph.D., Psychological Laboratory, Downing Street, Cambridge.

Visiting Foreign Members

Professor R. O. Rouse, M.R.C., Applied Psychology Unit, 15 Chaucer Road, Cambridge.
Professor C. H. Graham, Department of Psychology, University College, Gower Street, London, W.C.1.
Professor G. H. Bower, Department of Psychology, University College, Gower Street, London, W.C.1.

PUBLICATIONS RECEIVED

JOURNALS

Journal of Psychology. Edited by Mohammad Ajmal. Vol. No. 1, Dec., 1964. Issued by Department of Psychology, Government College, West Pakistan. Price Rs. 6s.

Psicologia y Educacion. Edited by Eloina Diaz Miniet and Louis Jones. Vol. 1, No. 2, April–June, 1964. Published quarterly by Department of Educational Psychology, Ministry of Education, Republic of Cuba. Annual subscription $5.00.

Acta Instituti Psychologici. Edited by Zoran Bujas. Nos. 22–48, 1961–64. Psiholojski Institut Sveucilista, Filozofski Fakultet, Zagreb. (Papers in English or French.)

Mental Health Book Review Index. Vol. 10, 1965. With an Editorial: In search of an organizing principle for the behavioral science literature. Pp. xx. + 78. Published by the Council on Research in Bibliography Inc., c/o Research Center for Mental Health, New York University, New York, 10003. Sponsored by the World Federation for Mental Health. International Council of Psychologists and Research Center for Mental Health, New York University. Annual subscription $5.00. Mental Health Book Review Index, Paul Klapper Library, Queens College, Flushing, New York, 11367, U.S.A.

MONOGRAPHS

Social Research and a National Policy for Science. Tavistock Pamphlet No. 7. London: Tavistock Institute of Human Relations. 1964. Pp. iv + 44. 7s.

Basic Analysis of Inner Psychological Functions. By E. Szekely. *Brit. J. Psychol. Monogr. No. 37*. London: Cambridge University Press. 1965. Pp. 141. 32s. 6d.

Nonparametric Trend Analysis: A Practical Guide for Research Workers. By George Andrew Ferguson. Montreal: McGill University Press 1965. Pp. v + 61. $1.95.

On Binocular Rivalry. By W. J. M. Levelt. Institute for Perception RVO-TNO, Soesterberg. 1965. Pp. vi + 110.

Sigmund Freud in Übersetzungen: Eine Bibliographie Seiner in Buchform und Anthologien Erschienenen Übersetzungen, 1945–1960/1. By Hans W. Bentz. Frankfurt: Hans W. Bentz Verlag. 1961. Pp. 60. DM 16.80.

BOOKS

Review in a later issue is not precluded by notice here

Progress in Experimental Personality Research. Vol. I. Edited by Brendan A. Maher. New York and London: Academic Press. 1964. Pp. x + 368. 76s.

Canine Behavior. By M. W. Fox. Springfield, Illinois: Thomas. 1964. Pp. xiii + 137. $6.75.

Principles of Treatment of Psychosomatic Disorders: Proceedings of a Conference held by the Society for Psychosomatic Research in London. Edited by Philip Hopkins and Heinz Wolff. Oxford and London: Pergamon. 1965. Pp. x + 118. 35s.

Psychiatrie Animale. Edited by A. Brion and Henri Ey. Paris: Desclée de Brouwer. Bibliotheque Neuro-Psychiatrique de Langue Française. 1964. Pp. 605. 480FB.

The Role of Psychosomatic Disorder in Adult Life: Proceedings of a Conference held by the Society for Psychosomatic Research in London. Edited by J. O. Wisdom and Heinz Wolff. Oxford and London: Pergamon. 1965. Pp. xiv + 229. 60s.

Psychosomatic Disorders in Adolescents and Young Adults: Proceedings of a Conference held by the Society for Psychosomatic Research in London. Edited by J. Hambling and Philip Hopkins. Oxford and London: Pergamon. 1965. Pp. ix + 246. 60s.

Hypnosis Induction Technics. By Myron Teitelbaum. With a Foreword by Michael M. Gilbert. Springfield, Illinois: Thomas. 1965. Pp. xiv + 184. $6.75.

Towards Self-Understanding: Group Techniques in Self-Confrontation. By Daniel I. Malamud and Solomon Machover. Springfield, Illinois: Thomas. 1965. Pp. xii + 279. $8.50.

Transference and Trial Adaptation. By Joost A. M. Meerloo and Marie Coleman Nelson. Springfield, Illinois: Thomas. 1965. Pp. vii + 155. $6.50.

Variables in Personality Theory and Personality Testing. An Interpretation. By Robert M. Allen. Springfield, Illinois: Thomas. 1965. Pp. ix + 96. $5.50.

Readings in Mathematical Psychology Vol. II. Edited by R. Duncan Luce, Robert Bush and Eugene Galanter. London and New York: Wiley. 1965. Pp. ix + 568. 68s.

Introduction to Developmental Psychiatry. By Beulah C. Rosselman, Ira M. Rosenthal and Marvin Schwarz. Springfield, Illinois: Thomas. 1965. Pp. v + 139. $6.50.

Speech Analysis, Synthesis and Perception. By J. L. Flanagan. Kommunikation und Kybernetik in Einzeldarstellungen, Band 3. Berlin, Heidelberg and New York: Springer-Verlag. (Distributors in U.S.A. and Canada—Academic Press, New York.) 1965. Pp. viii + 317. DM 58.

Compensatory Adaptations, Reflex Activity and the Brain. By Ezras A. Astratian. Translated from the Russian. International Series of Monographs on Cerebro-visceral and Behavioural Physiology and Conditioned Reflexes. Vol. I. London and Oxford: Pergamon. 1965. Pp. xv + 194. 80s.

Research in Behavior Modification: New Developments and Implications. By Leonard Krasner and Leonard P. Ullmann. New York: Holt, Rinehart & Winston. 1965. Pp. viii + 403. $8.95.

The Mentally Retarded Child: Development, Education and Treatment. By Max L. Hutt and Robert Gwyn Gibby. Boston: Allyn & Bacon. 1965. 2nd Edition. Pp. xiv + 441. $7.95.

Psychobiological Approaches to Social Behaviour. Edited by P. Herbert Leiderman and David Shapiro. London: Tavistock. 1965. Pp. xv + 203. 48s.

The Psychological Responses of Children to Hospitalization and Illness: A Review of the Literature. By David R. A. Vernon, Jeanne M. Foley, Raymond B. Sipowicz and Jerome L. Schulman. Springfield, Illinois: Thomas. 1965. Pp. vii + 192. $8.00.

Utopiates: The Use and Users of LSD 25. By Richard Blum and Associates. Foreword by Nevitt Sanford. London: Tavistock. 1965. Pp. xvi + 303. 58s.

Practice and Theory of Psychoanalysis. Vol. II. By Herman Nunberg. New York: International Universities Press. (London: Bailey Bros. & Swinfen.) 1965. Pp. 219. $4.50 (40s.).

The Group in Depth. By Helen E. Durkin. New York: International Universities Press. (London: Bailey Bros. & Swinfen.) 1965. Pp. xii + 378. $8.50 (76s.).

Medical Orthodoxy and the Future of Psychoanalysis. By K. R. Eissler. New York: International Universities Press. (London: Bailey Bros. & Swinfen.). 1965. Pp. x + 592. $12.00.

Personal and Organizational Change through Group Methods: The Laboratory Approach. By Edgar Schein and Warren G. Bennis. London and New York: Wiley. 1965. Pp. x + 376. 62s.

Drugs of Hallucination. By Sidney Cohen. Introduction by John Rowan Wilson. London: Secker & Warburg. 1965. Pp. xiii + 268. 30s.

Elementary Applied Statistics: for Students in Behavioural Science. By Linton C. Freeman. London and New York: Wiley. 1965. Pp. viii + 298. 53s.

Visual Capabilities in the Space Environment: A Collection of Articles Sponsored by the Human Factors Society of America. Edited by C. A. Baker. Oxford and London: Pergamon. 1965. Pp. vii + 203. 63s.

The Use of Written Communication in Psychotherapy. Edited by Leonard Pearson. Springfield, Illinois: Thomas. 1965. Pp. xiii + 65. $4.50.

Structural Models: An Introduction to the Theory of Directed Graphs. By Frank Harary, Robert Z. Norman and Dorwin Cartwright. London and New York: Wiley. 1965. Pp. ix + 415. 75s.

The Tactics of Psychotherapy: The Application of Psychoanalytic Theory to Psychotherapy. By William E. Murphy. New York: International Universities Press. (London: Bailey Bros. & Swinfen.) 1965. Pp. xiv + 623. $10.00 (90s.).

The Psychiatric Unit in a General Hospital: Its Current and Future Role. Edited by M. Ralph Kaufman New York: International Universities Press. (London: Bailey Bros. & Swinfen.) 1965. Pp. xvi + 482. $10.00 (90s.).

Sensory Restriction: Effects on Behavior. By Duane P. Schultz. New York and London: Academic Press. 1965. Pp. viii + 216. 60s.

Counseling in Marital and Sexual Problems: A Physician's Handbook. By Richard H. Klemer. Baltimore: Williams & Wilkins. (Edinburgh: Livingstone.) 1965. Pp. x + 309. 78s.

Determinants of Infant Behaviour III: Proceedings of the Tavistock Seminar on Mother-Infant Interaction. Edited by B. M. Foss. London: Methuen. 1965. Pp. xiii + 264. 50s.

PAPERBACKS

Hypnotherapy: A Survey of the Literature. By Margaret Brenman and Merton M. Gill. London and New York: Wiley. (The Menninger Foundation Series. No. 5. Science Edition.) 1964. Pp. vii + 276. 13s.

Visual Perception: The Nineteenth Century. By William Dember. London and New York: Wiley. (Perceptives in Psychology.) 1965. Pp. xii + 222. 23s. (cloth 38s.).

Insight Books. London: Van Nostrand. 1965.

- No. 24. *Personality and Science: A Search for Self-Awareness.* By Harold G. McCurdy. Pp. viii + 151. 12s.
- No. 25. *Science and Theory in Psychoanalysis.* Edited by Irwin G. Sarason. Pp. xiii + 205. 14s.
- No. 26. *Psychoanalysis and the Study of Behaviour.* Edited by Irwin G. Sarason. Pp. xiv + 173. 12s.
- No. 27. *Frustration and Conflict: Selected Readings.* Edited by Aubrey J. Yates. Pp. x + 246. 15s. 6d.
- No. 28. *The Experience of Psychotherapy: What It's Like for Client and Patient.* By William H. Fitts. Pp. iv + 188. 15s. 6d. ($1.95).
- No. 29. *Understanding Mental Health.* Edited by Robert I. Sutherland and Bert Kruger Smith. Pp. vi + 234. 15s. 6d. ($1.95).
- No. 30. *Cognitive Processes and the Brain.* Edited by Peter Milner and Stephen Glickman. Pp. viii + 216. 15s. 6d. ($1.95).
- No. 31. *Animal Drives.* Edited by George A. Cicala. Pp. vi + 266. 15s. 6d. ($1.95).

Principles of Animal Psychology. (Enlarged Edition.) By N. R. F. Maier and T. C. Schneirla. London: Constable. 1965. Pp. xvi + 683. 24s.

New Directions in Psychology II. By James and Marianne Olds, Frank Barron, William C. Dement, Ward Edwards, Harold Lindman and Lawrence D. Phillips. Foreword by Theodore M. Newcomb. New York: Holt, Rinehart & Winston. 1965. Pp. x + 422. $3.50.

A Manual of Laboratory Studies in Psychology. By Barbara Heckman and Robert Fried. London: Oxford University Press. 1965. Pp. xiii + 207. 20s

Laboratory Exercises in Psychology. By James N. Shafer. New York: Holt, Rinehart & Winston. 1965. Pp. 204. $2.95.

A Laboratory Manual for the Control and Analysis of Behavior. By Harlan Lane and Daryl Bem. Belmont, California: Wadsworth. 1965. Pp. xv + 283.

Freud: The Mind of the Moralist. By Philip Rieff. London: Methuen (University Paperbacks). 1965. Pp. xxiii + 397. 18s.

Laboratory Experience in Psychology: A First Term's Work. By B. Babington Smith. Oxford: Pergamon (The Commonwealth and International Library of Science Technology and Liberal Studies). Pp. xxiv + 240. 25s.

Perception and the External World. Readings selected, edited and introduced by R. J. Hirst. London and New York: Macmillan. 1965. Pp. viii + 310. 15s. ($1.95).

BOOK REVIEWS

Structure and Direction in Thinking. By D. E. Berlyne. London and New York: Wiley. 1965. Pp. xii + 378 68s.

Few psychologists are better equipped than Professor Berlyne to attempt the integration of the salient features of the distinctive contributions to the psychological study of thinking as they have been reported in the English, French and Russian languages in recent years. The approach which he adopts is confessedly neoassociationist and one which he says is usually identified with the term "stimulus-response (S-R) behaviour theory," although in his hands it exhibits a much more flexible and sophisticated guise than that usually evoked by the popular image of the S-R psychologist. Liberal use is made of concepts such as "implicit response" and "internal stimulus" on the grounds that the practice of classing central events as stimuli and responses must ultimately rest for its justification upon how fruitful it is in advancing knowledge a disarmingly reasonable plea, even if at times it leaves one with the sneaking feeling that one is being seduced by such "sweet reasonableness." From this neoassociationist starting point we are soon introduced to S-thinking, which is used to describe a kind of directed thinking in which responses are evoked by stimulus situations in which these would otherwise have been unlikely to occur and in which the result of the thinking process amounts to a transfer to one class of stimulus situations of a response pattern that is already associated with another class of situation. Likewise there is R-thinking, said to occur when a subject performs a response, or a combination of or sequence of responses that he has never before performed in any stimulus situation and that he would, in all likelihood, never have performed in any stimulus situation were it not for the thinking process.

This attempt to adhere to S-R type categories persists as Berlyne systematically examines a variety of contemporary and not-so-contemporary approaches to the study of thinking, through his neoassociationist eyes. In this way as information theory is dealt with, directed thinking becomes a means to information gathering and information rejection, and in considering theories built upon stimulus generalization we are told that the behavioural sequences resulting from directed thinking arise from secondary stimulus generalization. In this way habit-family hierarchies are viewed as special cases of mediated response generalizations wherein an equivalence exists between sequences of responses (behaviour chains) rather than between single responses. In turn Berlyne also considers the distinctions to be made between autistic and directed thinking, the relevance of Sechenov's theory of thinking, and of Piaget's conception of operation until we eventually reach what seems to be the core of the model of thinking that Berlyne wishes to present to us, namely, the functional importance of what he calls "transformational habit-family hierarchies." According to this formulation the traditional behaviourist account (following Watson, Hull and their successors) which depicts a train of thought as a sequence of implicit responses corresponding to successive events or stimulus situations, is modified so that these now relabelled "situational responses" together with their feedback stimuli become combined to form a "situational thought." Under the influence of Selz, Sechenov and Piaget, each successive pair of "situational responses" is linked by an intervening symbolic response called a "transformational response" which has its own feedback stimulus, the "transformational stimulus" and these two together comprise a *"transformational thought."* In this way the chains that constitute the habit-family hierarchies on which directed thinking depends will be composed of alternating situational thoughts and transformational thoughts.

Having thus introduced the building blocks of his model of directed thinking, Berlyne then shows how aspects of it had been partially foreshadowed by several earlier workers, including Tolman, Head, Bartlett, Spearman and, most recently, Newell, Shaw and Simon. Thus Bartlett's insistence in 1932 upon the importance of "effort after meaning" in recall foreshadows Berlyne's insistence that, "the recollection of something with a meaning that we understand depends primarily on retention of transformational responses" and that "if we have a starting point and series of transformations we do not need to store detail." In addition to chapters devoted to a more detailed exposition of some of the properties of his transformational chains and of the transformational hierarchies which these make up, Berlyne also give us a most penetrating discussion of the potential importance of habit-family hierarchies which possess structures corresponding to transitive (mathematical) groups. Such structures he believes occupy a unique position in the most

advanced forms of thinking and thus warrant careful experimentation investigation. In his discussion of the motivation, initiation and dynamics of directed thinking whilst we are not surprised to find "conflict" given such a prominent role, we may nonetheless wonder whether as a concept it is not so overworked that it is in danger of losing some of its usefulness.

This book is undoubtedly an important one, not only because it acquaints those who do not read Russian with literature hitherto not available in the West, but also because it accentuates the advantages and disadvantages of pushing an S-R approach to (and some may feel beyond) its limits. Certainly we shall not all agree with Berlyne's manner of linking up experimental findings and theorizing by S-R psychologists with our meagre knowledge and theorizing about complex thinking. However, if the more speculative parts of his theorizing do no more than goad us into a more sustained and systematic attempt to accumulate experimental results on high level thinking they will more than have justified their publication. In this respect even if we should concur with his own modest evaluation that "Many of the alternative conceptions of reasoning that have been proposed, including perhaps the one worked out in this book, are also open to criticism that they are inadequate as explanations and constitute little more than descriptions" (p. 335), we must at the same time agree that the description set out in this book does meet Berlyne's own criterion of a good description, which he says "should bring into prominence the variables that are of most moment and that are therefore likely to repay experimental study most bountifully" (p. 336). This it certainly does most elegantly and in a manner likely to stimulate fresh empirical research on thinking.

M. A. JEEVES.

Depression: A Cambridge Postgraduate Medical Course. Edited by E. Beresford Davies. London: Cambridge University Press. 1964. Pp. xv + 378. 90s.

Depressive States: A Pharmacotherapeutic Study. By Anthony Hordern. In collaboration with C. G. Burt and N. F. Holt. Springfield, Illinois: Thomas. 1965. Pp. xv + 166.

The value of publishing the details of discussion in symposium proceedings has often been queried. At first sight the advantages seem less obvious when the publishing delay reaches five years as it does in the case of the first of these two books. The long delay however provides an unexpected but sad pleasure. That of reading fresh and vivid contributions from the now silenced minds of Professor Kennedy and Professor Mayer-Gross. It is the more difficult to understand therefore why Mayer-Gross's paper should be printed only in abstract and without the references to the unusual yet important observations he was so expert at culling from the literature. Much of the discussion is disconnected. Decisions about where to cut must have been difficult but the reviewer has the strong impression that this failing is not that of the editor but rather reflects the fact that many participants were contributors manqué. At times it appears that they are concentrating more on the notes they have in their pockets than on relating their remarks to those of the previous speaker.

The standard of the major papers is uneven but many are well worth publishing and the book as a whole forms a useful picture of the different varieties of psychiatric thought. As a record of psychiatrists interacting at a time when the conflict between psychological and biochemical approach did not seem so unrealistic as it seems now, it is an invaluable social document.

The second book provides an instructive contrast. Considering the space into which it is compressed, the review of current views and research in the field of depression is comprehensive and clear. Brevity must inevitably lead to the superficial treatment of some aspects of the problem but there are plenty of leads to follow up in the extensive bibliography. However it is perhaps premature to see a scientific revolution in the double blind drug trial. As Professor Cade says in his introduction an impression of objectivity may be created by "meticulous methodology; exhaustive symptom analysis; that bulwark against observer bias, the double blind; adequate follow-up time; and expert statistical evaluation." It is a pity that in spite of these safeguards, the conclusions derived from this particular trial must be regarded as invalid. The flaw which invalidates the central conclusion, namely that amitriptyline is a more effective anti-depressant than imipramine, is that the equivalence of the dosage of the two drugs was determined by equivalence in weight. Apart from the fact that the molecular weight of imipramine is greater than that of amitriptyline the value of a drug is rarely a function of its efficacy per gram or per

molecule but is primarily a function of its efficacy at the maximum dose which is permissible in the light of toxic and side effects. Certainly the dose of imipramine used in the clinical trial reported in this book is less than that required in many depressive illnesses. As Scheff has pointed out, it is a function of the physician that he should be less concerned that his patient should suffer from over-investigation or over-treatment than that he should die because of under-investigation. The reverse motivation applies to the pharmaceutical industry. It is a far more valuable thing that many patients should remain unwell from under-treatment than that an enthusiastic coroner should have the opportunity to ascribe an unexplained death to a drug. Other minor criticisms of the therapeutic trial could be made but these would not diminish the value of this otherwise excellent source book on depressive illness.

Both the books reviewed here suffer from publication delays. That these delays are not necessary and unavoidable is indicated by the ability of May and Baker to publish in April the proceedings of the Leeds symposium on behavioural disorders which took place on 25th to 27th of the preceding March. This was admittedly a *tour de force*. However, the impact of being able to read what one's colleagues were thinking last week and to know what they are doing now is not only stimulating but inevitably must save a great deal of duplication of research effort.

A. ELITHORN.

The Approval Motive: Studies in Evaluative Dependence. By Douglas P. Crowne and David Marlowe. London and New York: Wiley. 1965. Pp. xii + 233. 57s.

The discovery of response sets has knocked the bottom out of most existing questionnaires. Crowne and Marlowe have done a lot of the work on the "social desirability" (S.D.) set—the tendency to show oneself in a favourable light. (The other important response set is "acquiescence"—the tendency to say "yes", or to "agree"). It is possible to predict a person's score on many questionnaires merely from knowing his response sets: the MMPI has been particularly hard hit—some of its scales correlate over 0·7 with S.D. Questionnaires *can* be constructed to avoid these sets—they should consist of multiple-choice items and the choices should be equated for S.D. Crowne and Marlow have devised a new scale for S.D., consisting of items such as "I never hesitate to go out of my way to help someone in trouble". They are interested in the personality characteristics of people who are high on S.D., and postulate that such people have a strong "need for approval". From this it is predicted that they should be more open to social influences. A series of studies is reported in which the top and bottom quartiles on the S.D. scale ($n = 30$, for each) are compared in a number of classical social influence situations. It is found that high scorers conform more, and show more rapid operant verbal conditioning, in a variety of experimental settings. The differences found between the experimental groups are large and consistent. One interesting result is that many of the high S.D. subjects show a higher threshold for obscene words. Evidence is presented to show that this is due to response withholding rather than to unconscious perceptual defence.

A further series of experiments is alleged to show that people high in need for approval avoid anticipated threats to self-esteem, and are defensive because they are vulnerable to reactions from others. These conclusions do not follow from the experimental results, without a great deal of intepretation, and other recent work on self-esteem and insecurity suggests that this is a much more complex problem. Nevertheless there are a number of quite interesting findings here—people high in need for approval finish therapy earlier, with less improvement; they were more readily aroused in the Schachter euphoria situation, and they were less popular than subjects who were less keen to be approved of.

Experimental social psychologists often ignore individual differences. Crowne and Marlowe have added to our knowledge by exploring one dimension of personality. It is unfortunate that they have not succeeded in throwing very much light on the psychology of this dimension.

MICHAEL ARGYLE.

The Pathology of Thinking. By Blyuma Vul'fovna Zeigarnik. The International Behavioral Sciences Series. Edited by Joseph Wortis. New York: Consultants Bureau. 1965. Pp. xvi + 211. $12.50.

To most psychologists in this country, Zeigarnik is simply an "effect" remembered from the far-off days of Gestalt Psychology. In fact, she is very much a person, who has been engaged for many years in the analysis of thought disorder in abnormal mental states. Her methods, which are for the most part very simple, have obvious affinities with those

of Goldstein, and include an object-sorting test of the kind he and Eva Rothmann originally devised in their studies of aphasia. She also uses an ingenious test involving pictograms suggested by Luria, the explanation of proverbs and metaphors, a picture-arrangement test akin to that of Wechsler, and certain variants of the word-association experiment. These techniques are described fairly fully though no attempt is made to provide quantitative standards of normal performance.

The results, which are given for the most part qualitatively, are of considerable interest. A sustained attempt is made to analyse test performance in relation to the different aspects of normal thinking, e.g. generalization and abstraction, the logical course of thought, and its purposiveness. At the same time, an indication is given of the incidence, and to some extent the severity, of thought disorder in a variety of neuropsychiatric conditions, such as cerebrovascular disease, brain injury, schizophrenia and psychopathy. It cannot, however, be said that this analysis is particularly systematic or that it bears at all closely on aetiological or diagnostic issues.

This study illustrates well both the advantages and the hazards of descriptive clinical inquiry. While Western clinical psychologists will be shocked, not without some reason, by the lack of scores, norms and statistics, neurologists and psychiatrists will recognize in Dr. Zeigarnik's work much that is familiar to them from their own experience which they have lacked time, and perhaps also talent, to pursue. The link with general psychology is also closer than is perhaps usual. Although Dr. Zeigarnik is at times a little unfair to Western clinical psychology, which she regards as both scientifically and ideologically suspect, there is no doubt that she seeks to establish the study of thought disorder on a genuine experimental basis. In spite of its technical deficiencies, her work shows clearly how much is to be gained by the use of simple methods, appropriately geared to issues in normal psychology, yet suitably adapted to the subtlety and many-sidedness of human thought and its disorders.

This book has been excellently translated by Dr. Basil Haigh. Professor Alexander Luria contributes an appreciative Foreword.

O. L. ZANGWILL.

Brain Function. Vol. II. UCLA Forum in Medical Sciences. No. 2. Proceedings of the Second Conference, 1962: RNA and Brain Function, Memory and Learning. Edited by Mary A. B. Brazier. Berkeley and Los Angeles: University of California Press. 1964. Pp. xvi + 360. $10.00. (Agents: Cambridge University Press, 1965, 80s.)

This seems to have been a wishful Symposium if ever there was one, and like most wishful enterprises distinctly autistic in its fulfilment. The idea seems to have been to gather together everyone above a certain level of scientific distinction whose work in his own field might be thought to bear, however tenuously, on the theme of the Conference. With breathtaking rapidity we pass from the ultrastructure of neurones to the hippocampus, from the planarian to the Korsakoff syndrome, and from RNA to the Space Program. In consequence, the Proceedings are virtually unreviewable.

Contributions of particular interest to psychologists include E. R. John on learning in planaria, L. Weiskrantz on temporal lobe ablation in monkeys, and M. Victor on the amnestic syndrome in man. Amazingly, most of the participants seem to have had something worth saying in discussion of most of the papers, which might seem to lend support to the idea of a general factor in human intelligence. Unfortunately, it seems more than ever uncertain whether a general factor will emerge in the elucidation of biological memory. However enlightening to the participants (and in lesser degree readers who toil through their Proceedings) these inter-disciplinary conferences may be, it is in the last resort on the strength of the individual disciplines, and on the quality of their practitioners, that scientific progress depends.

O. L. ZANGWILL.

The Child. Edited by William Kossen. London and New York: Wiley. 1965. Pp. xii + 301. 38s. cloth; 23s. paper.

This book is one of Drs. Kessen and Mandler's "Perspectives in Psychology" series. Like so many psychology books nowadays, it comprises nibblings from over a dozen different writers—ranging, in this instance, from the early seventeenth century (Locke, on Rewards, Reputation and Curiosity) to the present day (Piaget, on the Growth of Thought). The selected writers include Rousseau, Darwin, Binet and Simon, J. B. Watson, Freud and Anthony Ashley Cooper, Seventh Earl of Shaftesbury. In addition to excerpts from books by these and others, Kessen gives a general introduction and a few explanatory notes, linking each pair of adjacent chapters.

The reader of such a compilation is bound to find something of interest but he is likely, too, to ask himself whether yet another arbitrary though eclectic collection of this kind achieves very much. It is true that the reviewer had not known that the collaborator of Binet (who died in 1911) lived until 1961. And it is intriguing to learn that whilst Watson, in 1928, recommended *re* children, "Never hug and kiss them, never let them sit in your lap," William Cadogan, in 1749, recommended "to tumble and toss [the child] about a good deal, play with it, and keep it in good humour." The book, however, is rather for the seeker of *hors d'oeuvres* than for one who enjoys a hearty meal.

A. W. HEIM.

Infantile Autism: The Syndrome and its Implications for a Neural Theory of Behaviour. By Bernard Rimland. London: Methuen 1965 Pp. xi + 282. 30s.

As viewed by an experimental psychologist, speculations about childhood psychosis, of which infantile autism is a sub-category, have only just begun to emerge from the realms of folklore and superstition. Of course, excellent and perceptive clinical descriptions have been available for some time, but in the absence of definite experimental data explanatory hypotheses have run riot. It is to Rimland's credit that he disposes of one of the most prominent of these—the theory of the psychogenic origins of the illness which seeks the cause for the abnormality of the child in parental treatment and attitudes. In its place, the author attempts to put a more biologically orientated theory. Though laudable as an attempt, Rimland's view of the function of the reticular formation, which he selects as the locus of impairment in infantile autism, is somewhat naive. Moreover, the factual support for his hypothesis is not much greater than that for a psychogenic theory, which he rightly rejects on the grounds of insufficient evidence. Although the book contains some interesting suggestions, the present need in this field is for precise information resulting from adequately controlled experiments, rather than yet more speculation.

BEATE HERMELIN.

Inside the Black Room. By Jack A. Vernon. London: The Souvenir Press. 1964. Pp. xvii + 203. 21s.

This book gives a clear and straightforward account of the Princeton work on human sensory deprivation, with which the author was closely associated. Although nothing very new emerges, it is a useful review at the semi-popular level and deserves a better title.

O. L. ZANGWILL.

Hypnotherapy: A Survey of the Literature. By Margaret Brenman and Merton M. Gill. The Menninger Foundation Monograph Series No. 5. London: Wiley (Science Editions). 1964. Pp. xi + 276. 13s.

This is a paperback reissue of a book first published in the States in 1944 and which appeared in an English edition in 1947. It contains a review of the history and therapeutic applications of hypnosis, four case studies from the Menninger Clinic, and an experimental study of Lewinian "tension-systems" in relation to certain hypnotic phenomena. Although the book is by no means without interest, it is doubtful whether it merits re-publication.

O. L. ZANGWILL.

Principles of Training. By D. H. Holding. London: Pergamon Press. 1965. Pp. xvi + 156. 17s. 6d.

This is an addition to the Commonwealth and International Library of Science, Technology, Engineering and Liberal Studies and contains an introduction by Professor G. P. Meredith. Although adequately referenced it is clearly not aimed at scholars and is too abstract to be of immediate value to professional instructors. More likely to benefit are sixth formers wondering about a career in psychology, and first year undergraduates.

A chapter on knowledge of results is more advanced than this, and provides a neatly summarized classification of types of K.O.R. Other chapters discuss forms of guidance, relying on animal and maze learning studies a little too heavily for comfort in a book on training, training aids and verbal instructions. The section which deals with factors affecting learning curves and transfer illustrates one of hazards of trying to generate training principles from learning theory. For instance, distributed practice is better than massed practice. But then Dr. Holding honestly and clearly describes the difficulties of determining in any specific situation how much practice constitutes massing and how much spacing. The fault is not the author's; the conceptual gap between theories of

learning and principles of training is as wide today as it ever has been. The exception is in the area of programmed instruction, and this is another excellent chapter for the lay audience. The essential differences between linear and branching programming are well described and exemplified.

The style of writing is clear—a little flat—and the illustrations well chosen. For a near-paper-back the book is expensive. It would perhaps be more suitable to send to a nephew than to put into the department library.

R. CONRAD.

Handwriting in Psychological Interpretations. By A. G. Holt. Springfield, Illinois: Thomas. 1965. Pp. xiii + 262. $10.50.

There is no limit to the ways in which personality may be studied. In his handwriting a man undoubtedly expresses something of himself, but though many systems of interpretation have been put forward, none have gained general acceptance.

The author of this book claims that his method "not only reveals the true nature of a writer but also the impression a person conveys on and his attitude towards his world, how others see him and what they expect of him." No explanation is given of method, but over 800 illustrations, mostly of minute graphic details, are named and interpreted. Some of the interpretations recall what other writers have said, all but a few are unpleasant. The work cannot be regarded as a serious contribution to the study of personality.

B. BABINGTON SMITH.

Advancing Psychological Science. Vol. 4. Research in Physiological Psychology. Edited by Richard R. Louttit. Belmont, California: Wadsworth. 1965. Pp. x + 141.

The first three volumes of this paper-back series of selected readings were briefly noticed in this *Journal*, 1965, **17**, 186. The present volume contains three sections, dealing with Thirst, Mechanisms of Memory and Intracranial Self-Stimulation, respectively. Of these, only the first, with extracts from Cannon, Bellows and Andersson, tells a coherent story. The second contains much too much speculation and can be said to do little more than pose the problem of memory in its physiological aspects. The third deals with a topic still too recent to enable its long-term importance to be assessed. It is a little difficult to see what principles govern the selection of topics or what advantages the student is supposed to gain from perusal of these pre-packaged pickings from the periodicals.

O. L. ZANGWILL.

Delinquency and Child Guidance. Selected Papers by August Aichhorn. Edited by O. Fleischmann, P. Kramer and H. Ross. New York: International Universities Press Inc.; (London: Bailey Bros. & Swinfen). 1965. Pp. 244. $5.00 (45s.).

This is a short selection of papers by the author by *Wayward Youth*, who was a pioneer in the application of Freudian ideas to education. Among the issues dealt with are the juvenile court, reward and punishment in education, and child guidance as envisaged from the psycho-analytical standpoint. Miss Anna Freud contributes a brief though appreciative Foreword.

O. L. ZANGWILL.

Psychology of the Child. By Robert I. Watson. London and New York: Wiley. Second Edition. 1965. Pp. ix + 635. 57s.

This is a well-balanced text written in the conventional manner. It undoubtedly brings together a great deal of material on, or relevant to, child development and betrays an awareness of psychological principles. But to a reviewer with little knowledge of the field it all seems a bit uninspired.

O. L ZANGWILL.

Sigmund The Unserene: A Tragedy in Three Acts. By Percival Bailey, M.D. With a Foreword by Roy R. Grinker. Springfield, Illinois: Thomas. 1965. Pp. xvii + 127. $5.75.

Three semi-biographical and highly critical lectures about Freud, which are, however, a good deal better than the abysmal title might suggest. The author, who is a distinguished neurologist, is evidently the kind of person who can't take Freud and can't leave him alone. (His old friend and colleague, Dr. R. R. Grinker, who describes himself in his Foreword as a "psychoanalytic psychiatrist," has a shrewd idea why not.) Although Dr. Bailey's drama is unlikely to topple the great man from his pedestal, it might cut an inch or two from the stature of his official biographer, which would be no bad thing.

O. L. ZANGWILL.

Theories of Personality: Primary Sources and Research. Edited by Gardner Lindzey and Calvin S. Hall. London and New York: Wiley. 1965. Pp. xiii + 543. 53s.

This is a kind of scrapbook of personality research, organized in terms of contemporary theoretical approaches. It is divided into twelve sections, of which three are concerned with psychoanalysis and its derivatives and seven with writings by, or inspired by, Murray, Lewin, Goldstein, Allport, Sheldon, Rogers and Murphy. The two remaining sections are devoted to factor theories and stimulus-response theory respectively. Whereas many of the individual chapters have merit, the whole is overlong, indigestible and dispiriting. The only consolation is that 543 closely printed double-column pages at a penny farthing a page is, by prevailing standards, a decided bargain. O. L. ZANGWILL.

Colour Experiments with Modern Sources of Illumination. By K. v. Fieandt, Lea Ahonen, J. Järvinen and Arild Lian. Finnish Academy of Science and Letters, Series B, No. 134,2. Helsinki: Suomalainen Tiedadatemia. 1964. Pp. 89. 7.50 F.mk.

This excellent volume reports four interesting researches on the effects of fluorescent lighting on colour perception. Professor von Fieandt contributes a useful introduction, in which he stresses the seriousness of the problems created by modern lighting of the types normally called fluorescent and often giving apparently or nearly white illumination. The deception created by these light sources is, of course, greatest when they are most nearly white in appearance, and when their spectral energy distributions are widely different from that of normal white light. When they are manifestly yellow, as in sodium, or blue green, as in mercury lighting, the dangers of errors in colour judgment are still considerable, but not so great. Colour constancy depends on the presence of cues of the "real" colours, that is, the colours of an object in normal white illumination. If there are no such cues, and the illumination is exclusively in a light with energy distributions in the spectrum widely different from that of normal white light, even if it looks white, colour constancy will be limited to the influence of memory colours, to the knowledge of the colours certain objects normally have in true white light, and to perception of the illumination colour. Apart from such cues the object's colour will be dependent on the stimulus qualities, namely the light reflected from it in a given illumination. A dark blue motor car will look black in sodium lighting, unless there is the slightest bit of white or incandescent light shining on it, andthen it will look more like its real blue colour, but in fluorescent light overloaded with short waves it might look violet.

Dr. A. Lian contributes a useful study of chromatic contrast effects using different types of illumination. Professor von Fieandt's research report is about thresholds in colour perimetry for hues in fluorescent and extreme incandescent illumination. Dr. L. Ahonen gives us an examination of the degree of invariance in different object colours when shown in alternative illuminations. Dr. J. Jarvinen contributes a study of relative depth localization of chromatic surfaces in alternating incandescent and fluorescent illumination. Lastly, Professor von Fieandt makes a summary and concluding remarks. These are excellent examples of experiments in colour. R. W. PICKFORD.

THE QUARTERLY JOURNAL OF
EXPERIMENTAL PSYCHOLOGY

Vol. XVIII 1966 Part 2

SUCCESSIVE REVERSALS INVOLVING TWO CUES

BY

N. S. SUTHERLAND

From the Laboratory of Experimental Psychology, Sussex University

Rats were trained on a visual discrimination problem with two relevant cues—brightness and orientation. They were then given eight reversals in succession on the same problem. After reversals 7 and 8 they were tested with each cue presented on its own to see how much they had learned about each. Individual animals tended to reverse the cue about which they learned more from reversals 7 to 8, so that animals that had learned reversal 7 mainly in terms of the brightness cue learned reversal 8 mainly in terms of the orientation cue and *vice versa*. The result provides further confirmation for a two-process model of discrimination learning in which one process is that of selective attention.

Recent evidence strongly suggests that discrimination learning in animals involves at least two separate processes (Mackintosh, 1965; Sutherland, 1964*a*). Sutherland (1959, 1964*b*) has suggested one such model, in which the two processes are conceptualized as learning which analysers to switch in and learning what responses to attach to the outputs of those analysers. It is assumed that animals can analyse stimuli along many different dimensions, and in learning a discrimination they have to learn to switch in the analyser or analysers that classify stimuli along the dimension or dimensions on which the discriminanda differ. Only when an analyser is switched in can animals learn to attach the correct responses to its outputs and the more strongly any one analyser is switched in the slower will learning of responses to the outputs of other analysers proceed (Sutherland and Holgate, 1966). One of the phenomena the model explains is the effect of over-training on reversal. In most circumstances overtrained animals reverse faster than animals trained to a criterion, and it is assumed that overtraining strengthens the correct analyser more than it strengthens the correct response attachments. Thus the overtrained animal can learn to switch the response attachments to the relevant analyser faster than the non-overtrained (Mackintosh, 1962).

So far the model has not been applied to successive reversals, but it is known that if mammals are given a series of reversals their performance normally improves over the series as a whole. At the outset of each reversal, animals make a series of errors and the strength of the relevant analyser will decrease: the analyser will reach its lowest level at a time when the response strengths attached to the outputs of the analyser are equal for the two possible responses. As soon as the response attachments are reversed, the animal will be reinforced for using the correct analyser again and it will again be strengthened. If in a given reversal the analyser fails to drop as far as in the previous reversal before the response attachments are reversed, then the next reversal will tend to be faster, and at the end of that reversal the analyser will be higher than at the end of the previous one, and therefore will not

drop as far during the next reversal. Sutherland has simulated by computer a mathematical model of this kind and obtained successively faster reversals in stat-rats. The stat-rats also took longer to learn their first reversal than they took to learn their original discrimination, an effect also found with real rats.

Although this is one possible explanation of the improvement in performance shown over successive reversals, it is unlikely to be the whole story since animals may eventually come to reverse in one trial. This suggests that switching the response attachments can itself be conditioned to the experience of non-reward, and this would involve a mechanism so far not built into the theory. If a switch in response attachments can be conditioned to non-reward, one may ask whether a switch in the analyser used could also be conditioned to this cue. In most reversal situations there is more than one cue present and therefore more than one analyser that the animal could be using to discriminate between the stimuli; thus position as a cue could presumably be detected in a number of different ways possibly involving different sensory modalities. If the stimuli can be discriminated in a number of different ways, then there are several methods by which an animal could solve successive reversal problems.

To illustrate this, let us consider the case of an animal that is trained originally to select a white horizontal rectangle and avoid a black vertical rectangle, and is then given successive reversals with this pair of stimuli. An animal could learn to reverse, by learning at each reversal either about the orientation or brightness of the stimuli, and it could reverse its responses to one cue only or reverse its responses to both at once. It might, however, learn to change the cue to which it attended in each reversal while not altering the response attachments to that cue. Thus it could learn to go to white when paying attention to brightness and to go to vertical when paying attention to orientation, and switch the cue to which it attended from one reversal to the next. It is certain that the outputs from one analyser can control which analyser is used next otherwise an animal could never learn to switch in an analyser in a given situation. For example, if a rat is learning a brightness discrimination in a jumping stand, it must learn to switch in a brightness analyser when it is placed on the stand; since the information that it is on the stand is provided by the outputs from other analysers, the use of the brightness analyser must itself be conditioned to these outputs from other analysers. If the use of a given analyser can be conditioned to a situation, it is possible that a switch from one analyser to another could be conditioned to the experience of non-reward in a situation. The present experiment was designed to test this possibility.

Method

Subjects

The subjects were 24 male hooded rats from the colony maintained at the Oxford University Institute of Experimental Psychology. They were about 4 months old at the start of the experiment, and were maintained on a 22·5 hr. feeding schedule from 14 days before the experiment began and throughout its duration.

Apparatus

Animals were trained in a modified Lashley jumping stand. The animals had to learn to jump from a Y shaped platform to a ledge below the window bearing the positive stimulus; if they jumped correctly, the window was unlocked and they could push it down to obtain food for 15 sec. in a well behind the stimuli. If they jumped incorrectly, they found the stimulus window locked and were removed from the ledge in front of it after 5 sec. All training was by the non-correction method. All animals were pretrained to jump to plain grey windows. Trials were spaced at about 5 min. intervals and 20 trials a day were given. Further details of the apparatus and training procedures used appear in Mackintosh (1963).

Experimental design

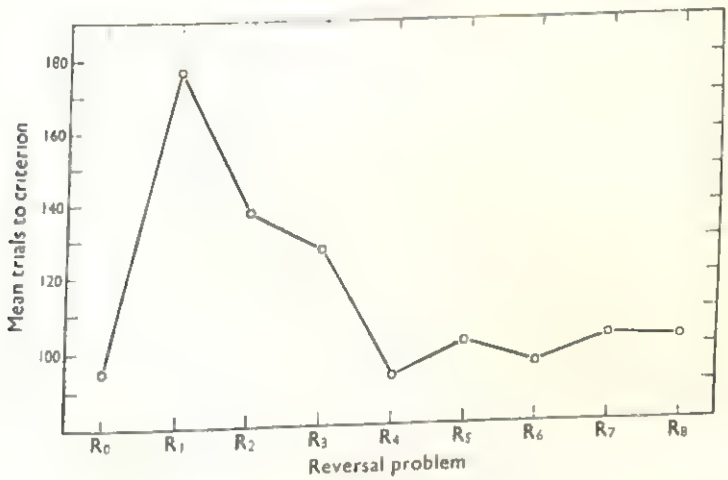
All animals were first trained (R_0) to jump to a white horizontal rectangle and avoid a black vertical rectangle. The rectangles were cut from ¼-in Perspex and mounted centrally on the grey stimulus windows: they measured 10 × 2·5 cm. When animals reached a criterion of 18 correct responses in any successive 20 trials, they were retrained (R_1) with the black vertical rectangle positive. They received eight reversals after their original training, and were taken to the same criterion on each. A non-correction method of training was used throughout. Following their 7th and 8th reversal (R_7 and R_8) each animal was given transfer tests alternating with retraining trials. On the retraining trials, animals were presented with the training shapes and rewarded for jumping to the shape which had been positive during the preceding reversal and not rewarded for jumping to the shape which had been negative during the preceding reversal. On transfer tests animals were rewarded whichever stimulus they chose. Ten transfer tests were given in a semi-random order with each of the following four pairs of shapes: white vs. black horizontal rectangles, white vs. black vertical rectangles, horizontal vs. vertical white rectangles, horizontal vs. vertical black rectangles. Performance on the first two pairs of shapes tests how much the animals had learned about the brightness cue during a given reversal, performance on the second two pairs tests how much they had learned about orientation.

RESULTS

Reversal learning

Figure 1 presents a graph of the number of trials to reach criterion during each successive reversal. The first reversal (R_1) took longer to learn than the original discrimination, but animals improved over the next three reversals.

FIGURE 1



Mean trials to reach criterion.

Transfer tests

Table I shows performance during the retraining trials and the transfer tests given after R_7 and R_8. There was no significant tendency for animals to perform better as a group on one cue than on the other in either set of transfer tests.

We can test in two ways for whether animals tend to leave the same responses attached to the orientation and the brightness analyser but to switch the analyser controlling the behaviour from one reversal to the next. (1) There should be a tendency for animals to reverse the cue on which they score more highly from one reversal to the next. Of the 24 subjects, 19 gave different scores both after R_7

and after R_8 on transfer tests with each cue. Of these animals, 14 reversed the cue on which they scored more highly between R_7 and R_8. The probability of this result being due to chance is less than 0·05 ($\chi^2 = 4{\cdot}26$). (2) A second test is to consider the changes in the scores on orientation and on brightness tests from R_7 to R_8. If the cue controlling behaviour changes from R_7 to R_8, we would expect that individual animals would show shifts in opposite directions for each cue between R_7 and R_8. For example if an animal scored less highly in orientation after R_7 than after R_8 we would expect it to score more highly on brightness after R_7 than after R_8. There were 21 animals that showed shifts in their scores on both orientation and on brightness tests between R_7 and R_8: of these 15 showed shifts in opposite directions on the two cues. Again the result is significant at better than the 5 per cent. level by a chi-squared test.

TABLE I

Per cent. Correct During Transfer

	Retraining trials (both cues)	*Transfer tests on brightness*	*Transfer tests on orientation*
R7	91·6	67·7	72·5
R8	91·0	73·8	69·6

Discussion

Although the effect was only significant at the 5 per cent. level, there was a tendency for animals to rely on different cues during R_7 and R_8 as evinced by the results of transfer tests. This suggests that animals can learn to switch from one analyser to another on successive reversals. The effect was not, however, an all or none one: thus, of the animals showing the effect, only one actually scored at chance (50 per cent.) or below on its less used cue over both sets of transfer tests given after R_7 and R_8. This suggests that most animals solved the reversal problems in two ways: they tended to learn predominantly about the outputs from opposite analysers on alternate reversals, but they still learned to change the response attachments to the weaker analyser on each reversal though not as effectively as they learned to change the response attachments to the stronger analyser. It must be remembered that switching from one analyser to another while leaving the response attachments the same on both is only one method of solution and that there is no reason to expect all animals to solve the problem in this way. The fact that a significant proportion of subjects gave evidence that they had solved the problem in this way can only be explained by a two process model of discrimination learning: the result cannot be explained unless animals can attach responses *separately* to each cue, i.e. unless there is some process of selective attention.

One other way in which the effect found might be brought about must be considered. It is known from other experiments (Sutherland and Holgate, 1966) that animals initially have a preference for the horizontal over the vertical rectangle and for black over white. If these preferences still showed after seven reversals, then we would expect that animals would do better on orientation than on brightness transfer tests after a reversal in which white horizontal was positive, and would do better on brightness transfer tests after a reversal in which black vertical was positive. If this hypothesis could explain the effect found, there would be no need

to assume that animals tended to switch the analyser about which they learned most from one reversal to the next. The data, however, refute the hypothesis. If it were correct, animals should tend to do better on brightness tests after R_7 than after R_8, since the preferred brightness value (black) was positive in R_7. Similarly they should tend to do better on orientation tests after R_8 than after R_7 since horizontal is preferred to vertical and horizontal was positive in R_8. Both the animals as a whole (see Table I) and those animals showing the effect of switching between analysers tended to do the reverse of this, performing better on brightness after R_8, and better on orientation after R_7 though the differences lack any statistical significance. The transfer test results were therefore not influenced by animals' initial preferences.

It may be that switching between analysers plays a part in the solution of many successive reversal problems. Thus even where the experimenter attempts to make the stimuli differ along only one dimension, there can be no guarantee that this difference is detected by only one analyser. For example the difference between a horizontal and a vertical rectangle could be analysed both by a mechanism that measured the amount of vertical and horizontal extent in the two shapes and by a mechanism that detected which stimulus extended lower in the visual field. Sutherland (1961) has obtained evidence that both these dimensions of difference between a horizontal and a vertical rectangle are analysed by the rat. Thus the present result may have implications for successive reversal problems even in cases where the experimenter used stimuli which at first sight appear to differ from one another only along one dimension.

One further finding merits brief discussion. In previous experiments (Sutherland and Holgate, 1966), it has been found that performance on retraining trials with two cues can be accurately predicted by performance on single cue transfer tests by applying the formula

$$P = p_1 \cdot p_2 \,/\, (p_1 \cdot p_2 + (1-p_1)\,(1-p_2))$$

where P is the probability of a correct response with two cues present and p_1 and p_2 are the probabilities of a correct response with each cue present on its own. When the formula is applied to the results of the transfer tests in the present experiment, it predicts 84·7 per cent. correct performance on the two cue retraining trials with both cues present after R_7 and 86·6 per cent. correct performance on retraining trials after R_8. Both figures considerably underestimate the scores actually obtained (*v.* Table I). It seems possible from the results of this experiment and from those of an earlier experiment (Sutherland, 1966) that after prolonged training, animals learn to make their responses not just to stimulus components but to compounds as well. Hence if the compound is not present as in the single cue tests there will be a severe fall-off in performance. Learning to base responses on compounds is of course compatible with learning about the components separately and with learning more about one component than the other. Thus switching in the dominant single cue analyser might be made to depend upon the cues having the correct compound values. This effect does, however, further complicate an already complex story.

This work forms part of a project on "Stimulus analysing mechanisms" supported by the American Office of Naval Research (Contract N62558-3927) and the Nuffield Foundation and undertaken at Oxford University.

The data analysis and writing up were undertaken whilst the author held a visiting appointment at M.I.T. This side of the work was supported in part by the Joint Services Electronics Program (Contract DA 36-039-AMCO3200(E)) by the U.S. Army

Research Office, Durham; and in part by the National Science Foundation (Grant GP-2495), the National Institutes of Health (Grant MH-04737-05), and the National Aeronautics and Space Administration (Grant NsG-496).

References

MACKINTOSH, N. J. (1962). The effects of overtraining on a reversal and a non-reversal switch. *J. comp. physiol. Psychol.*, **55**, 555–9.

MACKINTOSH, N. J. (1963). The effect of irrelevant cues on reversal learning in the rat. *Brit. J. Psychol.*, **54**, 127–34.

MACKINTOSH, N. J. (1965). Selective attention in animal discrimination learning. *Psychol. Bull.*, **64**, 124–50.

SUTHERLAND, N. S. (1959). Stimulus analysing mechanisms. In: *Proceedings of a Symposium on the Mechanization of Thought Processes*. Vol. II, 575–609. London: H.M.S.O.

SUTHERLAND, N. S. (1961). Visual discrimination of horizontal and vertical rectangles by rats on a new discrimination training apparatus. *Quart. J. exp. Psychol.*, **13**, 117–21.

SUTHERLAND, N. S. (1964*a*). Visual discrimination in animals. *Brit. Med. Bull.*, **20**, 54–9.

SUTHERLAND, N. S. (1964*b*). The learning of discrimination by animals. *Endeavour*, **23**, 148–52.

SUTHERLAND, N. S. (1966). Partial reinforcement and breadth of learning. *Quart. J. exp. Psychol.* (in press).

SUTHERLAND, N. S., and HOLGATE, V. (1966). Two-cue discrimination learning in rats. *J. comp. physiol. Psychol.* (in press).

Manuscript received 10th August, 1965.

REPETITION AND ALTERNATION IN RATS

BY

A. W. STILL*

From the Psychological Laboratory, Cambridge University

An experiment is reported in which rats were forced equally often to each arm of a T-maze, and, following each forced trial, were given a free trial. Group R were rewarded for repeating (i.e. choosing the side to which they had just been forced), group A for alternating. All rats began by alternating, but group R eventually learned to repeat. Various delays between free and forced trials were introduced, and the forgetting curve was found to be similar for the two groups. It is argued that these results show: (i) The decline in alternation with delay found in experiments on spontaneous and rewarded alternation is due to loss of information rather than a decline in the alternation tendency. (ii) The trace involved is sufficiently general to be regarded as a memory trace. The fact that rats can learn to repeat is inconsistent with the accounts of the memory trace suggested by the theories of Deutsch and Walker.

INTRODUCTION

Rats alternate spontaneously in a T-maze (Dember and Fowler, 1958) and delayed alternation, where a rat is rewarded for alternating, is a familiar delayed response task. Alternation, whether spontaneous or rewarded, involves retention of information, since choice depends upon which arm was last entered; thus some kind of memory trace must be laid down when a rat enters an arm, and alternation may be used as a test of memory. Previous experiments show that alternation declines gradually as the intertrial interval is increased up to several hours whether the alternation is rewarded (Ladieu, 1944; Petrinovitch and Bolles, 1957) or spontaneous (Zeaman and House, 1951; Walker, 1956); if these results are compared with conventional delayed response experiments, such as Smith (1951), in which performance declines to chance within a few seconds, the forgetting curves for rewarded and spontaneous alternation appear strikingly similar, which suggests that the same traces and forgetting processes are involved.

Before using alternation to investigate memory two preliminary questions must be answered:

(i) Is the decline in alternation with delay due to loss of information, or is it due to a decline in the alternation tendency itself?

(ii) Is the trace involved sufficiently general to be regarded as a memory trace? It is possible, for instance, that the information stored can only be retrieved in alternation, and cannot be used as the basis of any other response. This may be tested by rewarding a rat only if it repeats its previous choice; if it gives up alternating, and learns to repeat, this would indicate that other responses, besides alternation, can be based upon the trace. Furthermore, if the same trace is involved both in delayed alternation and delayed repetition, performance should be a similar function of delay in the two cases. If this is found to be so, then an answer is provided to the first question; for if the decline in performance with delay in rats rewarded for alternating is due to a decline in the alternation tendency, a similar decline would not be predicted in rats rewarded for repeating; while if it is due to loss of information the decline should be the same whether rats are rewarded for alternation or repetition.

* Now at Psychology Department, University of Durham.

Among existing theories of spontaneous alternation, at least two suggest an account of the trace involved in which it is not a memory trace in the general sense discussed above, and predict that it should not be possible to teach rats to repeat. The two theories are:

(1) *Deutsch* (1960). According to this theory, the probability of a rat's approaching a stimulus depends upon the amount of excitation in the "link" representing the stimulus. Excitation enters a chain of links at the link representing the reward stimuli, and is conducted down the chain to the choice point links. Spontaneous alternation occurs in rewarded rats because, when an arm is entered, its links become temporarily refractory to the passage of excitation; the memory trace is this refractoriness. In order to repeat, therefore, a rat must choose the arm whose links are refractory and have least excitation; but, *ex hypothesi*, it chooses the arm whose links have the most excitation, so it should not learn to repeat.

(2) *Walker* (1958). Part of the "action system" underlying an action goes through a process of "recycling" following the action. During recycling, potentiated phases alternate with decremental phases; since the probability of the action is reduced during the decremental phase, and this lasts for a relatively longer time than the potentiated phase, alternation between actions will tend to occur. The memory trace is the recycling and to learn to repeat, therefore, a rat must learn to carry out the action whose system is predominantly decremental; as with Deutsch, this conflicts with the principle on which choices are supposed to be based.

In the following experiment one group is rewarded for repeating, the other for alternating, and performance is tested as a function of delay.

Method

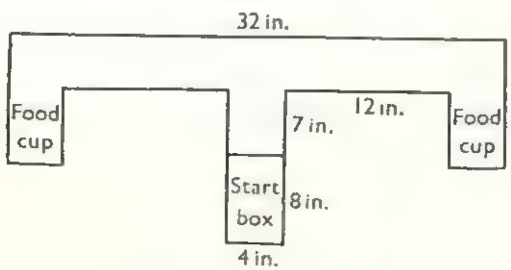
Subjects

Ten female hooded rats, about 6 months old at the beginning of the experiment, were obtained from N.I.M.R., Mill Hill. They had previously been used in a runway experiment (a repeat of the experiment by Chapman and Levy, 1957).

Apparatus

An alley maze in the shape of an E, as shown in Figure 1, was used. The alleys were 4 in. wide and 6 in. high, and the maze was of unpainted aluminium throughout, except

FIGURE 1



Floor plan of the maze used.

that the floor of the left arm was black, and of the right arm white, and the roof was of wire netting; lighting was provided by two 30-watt striplight bulbs suspended 18 in. above the roof of the maze, one on each side of the choice point. Guillotine doors controlled access to the arms of the maze from the stem, and these could be raised and lowered from behind the start box by means of string. At the ends of the arms of the maze were removable food cups, in which the reward was placed; they were raised ⅜ in. off the floor, and were ⅝ in. in diameter; the reward consisted of broken pieces of standard food pellets (Oxoid 41b), each piece weighing approximately 150 mg.

Procedure

Days 1–45. On all trials the rat was put into the start box and 5 sec. later the guillotine door leading from the start box to the stem was raised. The trials were given in pairs. On the first trial of a pair the rat was forced down one arm, being prevented from entering the other arm by means of a guillotine door at the choice point; it was rewarded by one piece of food in the food cup, and, after it had eaten this, was picked up and put back into an individual retaining cage, 6 ft. behind the start box of the maze.

After a delay a further trial was given, with both arms open. The rats were randomly divided into two equal groups:

Group R. On the second trial of a pair reward was only given if the rat repeated the response it had been forced to make on the preceding trial. If it alternated it found neither food nor food cup at the end of the arm.

Group A. On the second trial of a pair reward was only given if the rat alternated.

Following an error, the guillotine door to the chosen arm was closed behind the rat, which was then left in the maze for 15 sec.

Unless otherwise stated, two pairs of trials were given per day, one at about 10.00 a.m., and the other at 10.00 p.m. After each of these two daily sessions, when all the rats had been given their pair of trials, they were taken back into the keeping room and 5 min. later they were given food for 1 hr. They were thus about 11 hr. hungry when tested. For the first 10 days the intertrial interval was 20 sec.; for the next 23 days it was 15 min.; and for the final 12 days it was 20 sec. again. Pairs of trials were grouped into blocks of four and over each block of four pairs a rat was forced on the first trials alternately LRRL and RLLR (R = right, L = left). Three rats from each group began with one such sequence, two began with the other sequence. Any rat not entering the open arm on the first trial of a pair within 3 min. of being put into the maze was pushed into the arm with a closed fist (there was only one rat, from group R, that had to be pushed in this way, and this was only early on in training); on the second trial of a pair, rats were left in the maze until they had chosen.

After 45 days two of the rats from each group were discarded. In group R, only three rats showed signs of learning to repeat, and these were retained. In group A one rat was discarded for consistently biting, and the other rat discarded was the only rat of the group that failed to alternate above chance.

Days 46–56. The procedure was as before, except that delays between trials were alternately 15 min. and 25 sec.; an exception to this was that the evening pair of trials on day 50 had the same delay as the morning trials on day 52; this ensured that the two delays were given equally often in the morning and evening.

Days 57–60. The delay between trials was increased to 30 min., otherwise the procedure was the same, except that the rats were fed only once per day, for 2 hr., after the morning trials; they were thus more highly motivated for the morning than the evening trials.

Days 61–64. The procedure was as for the previous 4 days, except that the delay was increased to 1 hr.

Days 65–68. The procedure was as for the previous 4 days, except that the rats were rewarded on only half the pre-delay trials. Half the rewarded pre-delay trials were given in the morning, half in the evening.

Days 69–70. The procedure was as for days 61–64, except that food was present in both arms at the time of choice. This provides a control against the possibility that the rats can smell the food from the choice point and choose on the basis of this cue.

Days 71–78. Delay was increased to 4 hr. and only one pair of trials was given per day. This was given in the morning and the rats were fed in the evening, so that they were about 10 hr. hungry at the time of the first daily trial.

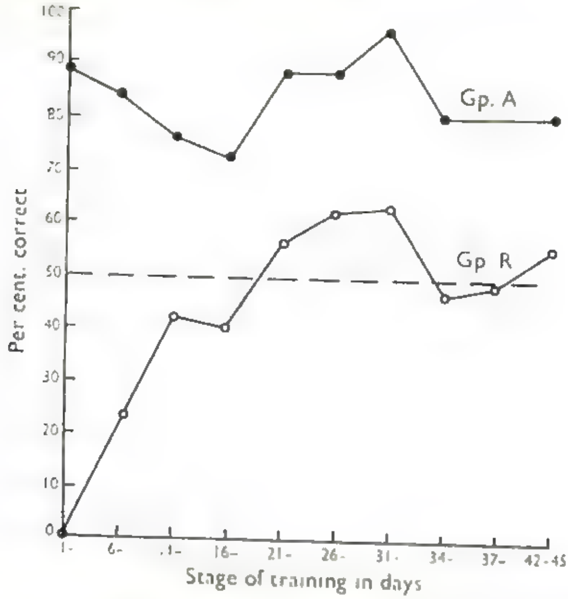
Days 79–90. The procedure was as for days 71–78, except that half the pre-delay trials were unrewarded.

Results

Days 1–45

Figure 2 shows the percentage correct for each group (excluding the rats discarded after day 45). Rats failing to learn, and discarded, showed strong position preferences after the first few trials.

FIGURE 2



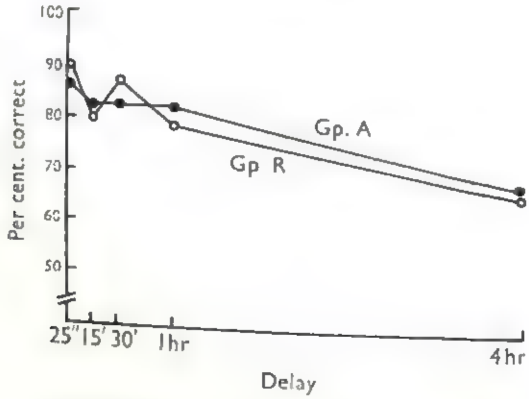
Performance as a function of stage of training, in rats rewarded for alternating (Group A) and repeating (Group R). Delay between trials was 20 sec. on days 1–10 and 34–45, and 15 min. on days 11–33.

After day 45

All the remaining rats in both groups show a high level of performance. By the binomial test (Siegel, 1956) the overall performance for each rat is significantly above chance at the 0·001 level.

Performance as a function of delay. Figure 3 shows the percentage correct for each group at each delay.

FIGURE 3



Performance as a function of delay.

Control trials for olfactory cues. On days 69–70, when olfactory cues could not aid the rats, the rats of group R repeated four, three and three times (out of four) respectively. The group percentage of correct choices is therefore 83 per cent.

Reward on the first trial. Table I shows performance as a function of reinforcement on the first trial for all days when first reinforcement is varied. Both groups appear to perform above chance, whether rewarded on the first trial or not. In both groups

TABLE I

Per cent. Correct as a Function of Reward on the Pre-Delay Trial; Days 65–68 and 79–90

	Reward pre-delay	*No reward pre-delay*
Group A	73·3	80·0
Group R	76·7	70·0

there is a slightly greater tendency to repeat following a rewarded trial, so that group A performs less well following a rewarded trial, while group R performs better. However, this tendency is not significant even when the groups are combined, since two rats, one from each group, perform at the same rate under the two conditions, leaving four rats tending to repeat more when the first trial of a pair is rewarded (by the signs test, one-tailed $p = 0{\cdot}0625$).

Discussion

It appears that at least some rats can learn to repeat an immediately preceding response, though it would be rash to conclude from Figure 3 that forgetting curves are the same whether rats are rewarded for repeating or alternating; for, (*a*) rats were discarded from the two groups on different grounds; (*b*) the numbers in the groups were very small. Nevertheless, relative to the results of conventional delayed response experiments (such as that of Smith, 1951), it is clear that the forgetting curves are very similar to each other and also to the curves from other delayed and spontaneous alternation experiments; this suggests that the same memory trace and forgetting processes are involved. If so, we can conclude that the trace involved in alternation can form the basis of responses other than alternation, and that (i) the decline in alternation with delay is due to loss of information; (ii) the trace is sufficiently general to be regarded as a memory trace.

The results conflict with the account of the memory trace suggested by the theories of Deutsch (1960) and Walker (1958). Thus, while these theories may still apply to some of the areas of behaviour for which they were intended, we must look elsewhere for a satisfactory account of the memory trace in spontaneous and delayed alternation.

The author would like to thank the M.R.C. for financial support; Mr. A. J. Watson, who supervised this research, for much helpful discussion and advice; and Professor O. L. Zangwill for his continued encouragement.

References

Chapman, R. M., and Levy, N. (1957). Hunger drive and reinforcing effect of novel stimuli. *J. comp. physiol. Psychol.*, **50**, 233–8.

Dember, W. N., and Fowler, H. (1958). Spontaneous alternation behaviour. *Psychol. Bull.*, **55**, 412–28.

Deutsch, J. A. (1960). *The Structural Basis of Behaviour*. London: Cambridge University Press.

LADIEU, G. (1944). The effect of length of delay interval upon delayed alternation in the albino rat. *J. comp. Psychol.*, **37**, 273–86.

PETRINOVITCH, L., and BOLLES, R. (1957). Delayed alternation: Evidence for symbolic processes in the rat. *J. comp. physiol Psychol.*, **50**, 363–5.

SIEGEL, S. (1956). *Non-parametric Statistics for the Behavioural Sciences.* New York: McGraw-Hill.

SMITH, M. P. (1951). The stimulus trace gradient in visual discrimination learning. *J. comp. physiol. Psychol.*, **44**, 154–61.

WALKER, E. L. (1958). Action decrement and its relation to learning *Psychol. Rev.*, influence of reward. *J. comp. physiol. Psychol.*, **49**, 167–76.

WALKER, E. L. (1958). Action decrement and its relation to learning. *Psychol. Rev.*, **65**, 129–42.

ZEAMAN, D., and HOUSE, B. J. (1951). The growth and decay of reactive inhibition as measured by alternation behaviour. *J. exp. Psychol.*, **41**, 177–86.

Manuscript received 9th October, 1965.

SPONTANEOUS ALTERNATION AND PATTERN OF REINFORCEMENT

BY

A. W. STILL*

From the Psychological Laboratory, Cambridge University

In Experiment I, rats were given one pair of trials per day in a T-maze; group RN were rewarded on the first trial only, NR were rewarded on the second trial only, while RR were rewarded on both trials. Alternation was approximately the same for all groups, and there was little decline in alternation over 28 days. There was, however, a difference in the choice behaviour on trials when rats failed to alternate, NR and RR, but not RN, choosing according to a position preference. These results are incompatible with several earlier theories of spontaneous alternation, and the hypothesis was put forward that alternation is initially part of an exploratory tendency, but that, when the maze becomes explored, the exploratory tendency dies down, and alternation only continues when rats have been reinforced (primarily or secondarily) for alternating. This hypothesis was tested in a second experiment, similar to the first but in which one group of rats was not rewarded on either trial; as predicted, this group showed significantly more decline in alternation than rewarded rats.

Introduction

It has been found (Still, 1966) that rats can learn to repeat their choices, which is difficult to explain on the theories of spontaneous alternation put forward by Deutsch (1960) and Walker (1958). In the present paper further predictions from these and another theory are tested in an experiment in which rats are given one pair of trials per day; for one group the first trial is rewarded, for another the second, and for the third group both trials are rewarded. The theories tested are:

(1) *Deutsch* (1960). Choice in a T-maze depends upon the relative excitation in the links representing the choice point stimuli of the two arms. Excitation flows down the chains of links from the link corresponding to the reward, and where both arms are equally rewarded the excitation at the choice point will depend upon the resistance to excitation in the chains of links. This resistance increases (*a*) following an entry to an arm; (*b*) when a stimulus that usually appears fails to do so, as when reward is withheld; the increase due to (*b*), which underlies extinction, lasts longer than that due to (*a*). When reward is withheld there will thus be two factors making for increased resistance, while when the rat is rewarded there will only be (*a*); so resistance, and therefore probability of alternation, should be greater following unrewarded trials. Furthermore, the difference between the two conditions will increase with the delay between trials, since the only source of alternation when the rat is rewarded is relatively temporary.

Prediction. Rats rewarded on the first trial of a pair should show less alternation than rats not rewarded on the first trial, provided that they are at least sometimes rewarded. The difference should increase with delay.

(2) *Walker* (1958). An action gives rise to recycling in the action system underlying the action; if the action is reinforced, this recycling process will initially be more vigorous than when it is not reinforced, but it will decline more rapidly, and leave the action system at a lower threshold level; the more vigorous the

* Now at Psychology Department, University of Durham.

recycling, the greater the probability of alternation, so that, immediately after an action (e.g. a choice to the left or right in a T-maze) alternation will be greater if the action was rewarded. With a longer delay, however, alternation will be more likely following an unrewarded trial, since the recycling process dies down more slowly.

Prediction. Rats rewarded on the first trial of a pair should show *more* alternation than rats not rewarded on the first trial. This difference should decrease with delay, and eventually reverse.

(3) *Montgomery* (1951). Alternation is assumed to occur initially as a result of "exploratory tendencies," which decline in strength as the maze is explored. In spite of this decline, alternation may continue as a learned response to proprioceptive stimulus traces set up by the previous response; these traces are estimated to have faded away by about 90 sec.

Prediction. Alternation should decline with training if the second trial of a pair is unrewarded, but not if it is rewarded, except when the delay between trials is over 90 sec.

In most previous experiments using rats rewarded on all trials alternation has not been found to decline with training (Wingfield and Dennis, 1934; Zeaman and House, 1951; Rothkopf and Zeaman, 1952). Weitz and Wakeman (1941) did find a decline, but this is probably related to the fact that they used a common goal box, since Sutherland (1957) found more decline in a maze with a common goal box, than in one with separate goal boxes.

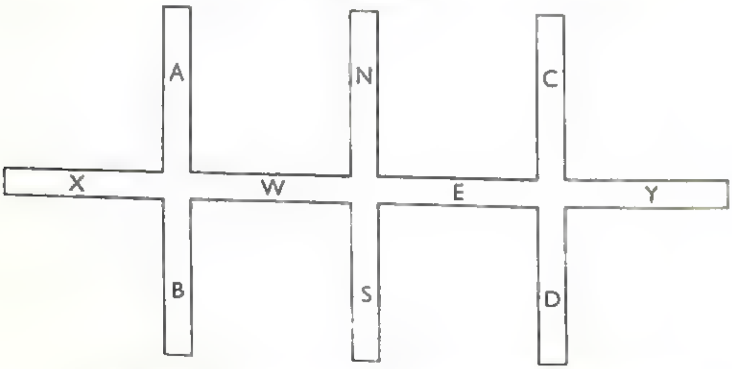
Experiment I

Method

Subjects. Twenty-four male hooded rats, about 3 months old at the beginning of the experiment, were obtained from N.I.M.R., Mill Hill, London.

Apparatus. An alley maze was used, as shown in Figure 1. For the actual experiment only alleys W, E and S were used, forming a T-maze divided off from the rest of the maze by means of aluminium doors. Some of the other alleys were used in the pre-training, described below. The inside height of the alleys was 5 in., and the maze was of unpainted aluminium throughout, except that the walls of alley W were painted white, of alley E black; the roofs were of Perspex. Lighting was provided by three 30-watt strip-light bulbs, 18 in. above the roofs of W, E and S. Guillotine doors controlled access to the arms W and E from the stem. At the ends of each arm were food holes, 2 in. in diameter, which could be closed by means of sliding doors; these doors were flush with, and just below, the floor of the maze. The holes contained Skinner box pellets, made from standard laboratory diet.

FIGURE 1



Floor plan of maze used. For description see text.

Pre-training

Days 1–4. For 5 min. per day rats were handled and allowed to explore a flat aluminium platform (1 ft. × 3 ft.) in groups of four. They were satiated for food and water.

Days 5–8. For 5 min. per day the rats were allowed to explore one of the T-mazes ABX, CDY of Figure 1, in groups of four. They were satiated for food and water, and the food holes in the maze were closed. Each rat experienced each of the two T-mazes equally.

Days 9–12. For 5 min. per day each rat was allowed to explore, on its own, one of the arms of the T-mazes used on days 5–8. After 2 min., the food hole at one end of the arm was opened, and feeding was permitted for the remaining 3 min. The rats were hungry, having food in the living cages for 3 hr. per day, beginning 10 min. after each day's training session; this schedule was kept to throughout the experiment, except where otherwise stated.

Days 13–16. The rats were trained to run along an alley-way, made up of the two arms of one of the alleyways used on days 5–8 (i.e. alley AB, or alley CD). They were given two trials per day; one food hole was open, and 15 sec eating allowed. The rats started at one end of the alley and ran to the other, the four possible starting points were given equally often to each rat.

Days 17–20. As for days 13–16 except that alley SN was used, the rats always being started from S. After the end of the training session on day 20, the rats were randomly divided into three groups of seven rats each, three rats being discarded for slowness in running and eating.

Days 21–22. A test of alternation with a long delay between trials was given. The rats were first given one forced trial in the experimental maze WES. Half were forced to one arm (the door to the other arm being closed) half to the other. The food holes were open, and 15 sec. eating allowed. 15 min later the rats were given food for 2 hr. in their living cages. $17\frac{1}{4}$ hr. after the forced trial one rewarded trial was given, which was free (both arms of the T-maze being open). 15 min. later the rats were given food in their living cages for 4 hr Previously (and subsequently until day 51) they were run in the morning. One rat run in the afternoon; for the free trial of day 22 they were run in the morning. One rat was discarded for failing to choose within 5 min. on day 22.

Testing

Days 23–50. The three groups (one now containing six rats, two containing seven) were randomly assigned to three conditions:

Group RN: Always rewarded on the first of the daily pair of trials, but not on the second.

Group RR: Rewarded on both trials (this group contained six rats).

Group NR: Rewarded on the second, but not the first trial.

After 10 days testing a rat from RR died, and therefore the results of only five rats from this group are considered.

Each rat was given one pair of trials per day in the T-maze WES. The first trial was forced equally often to each arm. Pairs of trials were grouped into blocks of four, and over each such block the rat was forced on the first trial in the order LRRL or RLLR (L = left, R = right), these two sequences being used over alternate blocks. Approximately half the rats from each group started with one of these sequences, half started with the other. The second trial of each pair was free. The intertrial interval was 25 sec., during which the rats were put into individual cages 5 ft. behind the stem, while the necessary adjustments were made to the maze. On rewarded trials, the food holes at the end of the arms were left open; on unrewarded trials they were closed. In either case the rat was left in the chosen arm, with the guillotine door to that arm closed behind it, for 15 sec. following the choice.

Days 51–72. Eleven different delays, varying between 5 sec. and 12 hr., were introduced between the trials of a daily pair. On any one day each rat was given the same delay. The order of delays over days 51–61 was random, with the restriction that each delay was given only once. Over days 62–72 the order was reversed. Thus each rat was given two pairs of trials at each delay; at each delay it was forced once to the left, and once to the right.

During these 22 days, the first trial of a day was given at about 10.00 a.m. and feeding begun at about 4.00 p.m., never less than 10 min. after the last trial of a pair. An exception, which obviously could not be fitted into this timetable, was the 12-hr

delay; for this, the first trial was begun just after 9.00 a.m., feeding was begun just after 9.00 p.m., and food was left in until midday the next day. There was therefore no testing on the day after the 12-hr. delay.

After day 61 there was a break of 11 days without testing.

Results

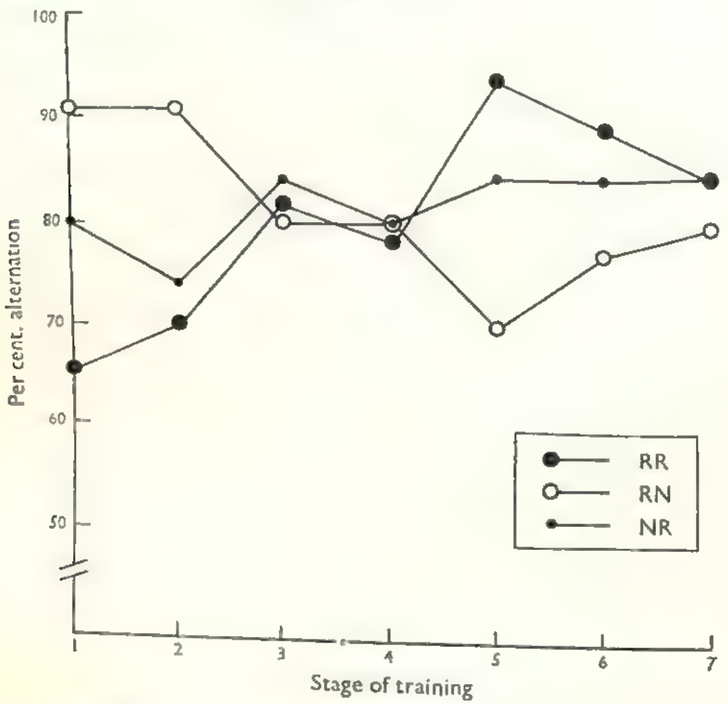
Days 21–22

Fifteen out of 20 rats alternated with $17\frac{1}{4}$ hr. between trials. This differs significantly from chance by the binomial test (Siegel, 1956); one-tailed $p = 0{\cdot}021$.

Days 23–50

Figure 2 shows percentage alternation for each group over each block of 4 days. Table I shows the total alternation for each rat and the sums weighted by orthogonal polynomials of the first order (Fisher, 1958). In order to compare groups both with

FIGURE 2



Per cent. alternation over each block of 4 days during days 23–50, for rats rewarded on the first daily trial only (RN), both daily trials (RR) and the second daily trial only (NR).

respect to total alternation and trend of alternation against training, total alternation and trend coefficients are compared by means of the Kruskal-Wallis non-parametric analysis of variance (Siegel, 1956), and by the Mann-Whitney U-test. The use of non-parametric tests in conjunction with orthogonal polynomials is explained elsewhere (Still, 1967).

Difference in total alternation. None of the differences between groups in total alternation is significant at the one-tailed 5 per cent. level.

Difference in trend of alternation against training. Comparing first order coefficients, Kruskal-Wallis H = 8·36 ($p < 0{\cdot}02$). RR differs significantly from both

TABLE I

TOTAL ALTERNATION BY EACH RAT (OUT OF 28) ON DAYS 23–50, AND SUMS WEIGHTED BY ORTHOGONAL POLYNOMIALS OF THE FIRST ORDER

Group RN			*Group RR*			*Group NR*		
Rat	*Total*	*Trend coefft.*	*Rat*	*Total*	*Trend coefft.*	*Rat*	*Total*	*Trend coefft.*
1	28	0	1	19	+10	1	28	0
2	18	−5	2	27	+3	2	15	−1
3	24	−6	3	24	+4	3	23	+8
4	24	−4	4	17	+2	4	14	0
5	24	−9	5	27	+3	5	28	0
6	25	+3				6	28	0
7	20	−1				7	27	+2
Mean percentage	83·2	Mean trend −2·6	Mean percentage	81·4	Mean trend = 4·4	Mean percentage	83·2	Mean trend = 1·3

RN (Mann-Whitney $U = 2$, one-tailed $p < 0{\cdot}005$) and NR ($U = 4\frac{1}{2}$, one-tailed $p < 0{\cdot}02$). Analysing the groups individually by means of Wilcoxon's T-test, only RR shows a significant linear trend ($T = 0$, one-tailed $p = 0{\cdot}031$); i.e. RR shows a significant increase in alternation with training. There are no significant differences due to higher order trends.

Analysis of trials where the rat fails to alternate. If a rat shows above chance alternation, a certain proportion of his choices are accounted for by the alternation tendency (whatever that may be due to), but, if there is less than 100 per cent. alternation, some of the choices must presumably be due to some other tendency (unless they are random). The rat might, for instance, choose according to a position preference; if this were so, then it would be manifested in a tendency for all a rat's choices, when it failed to alternate (i.e. on repetition trials), to be to the same arm of the maze. Suppose that on repetition trials a rat chooses arm W, w times, and arm E, e times, so that it fails to alternate $w + e$ times; then the expected number of choices to each arm on repetition trials will be $\frac{1}{2}(w + e)$. The value of $w - \frac{1}{2}(w + e)$ is an indication of the extent of the preference for W compared with E, and the closer this value is to zero, the more justified we are in concluding that there is not a preference for one arm over the other. This suggests that an index of degree of position preference on repetition trials may be obtained by expressing $w - \frac{1}{2}(w + e)$ in standard deviation units to make it independent of the actual number of such trials, i.e.:

$$I_{pref} = \frac{w - \frac{1}{2}(w + e)}{\sqrt{(w + e).\frac{1}{2}.\frac{1}{2}}} = \frac{w - e}{\sqrt{w + e}}$$

where I_{pref} is the index of degree of position preference. Values of this index for each rat, except those that always alternated, are shown in Table II. Comparing the groups, Kruskal-Wallis $H = 7{\cdot}62$ ($p < 0{\cdot}05$). RN is significantly less than RR (Mann-Whitney $U = 3$, two-tailed $p < 0{\cdot}05$) and than NR ($U = 1\frac{1}{2}$, two-tailed $p < 0{\cdot}05$).

Days 51–72

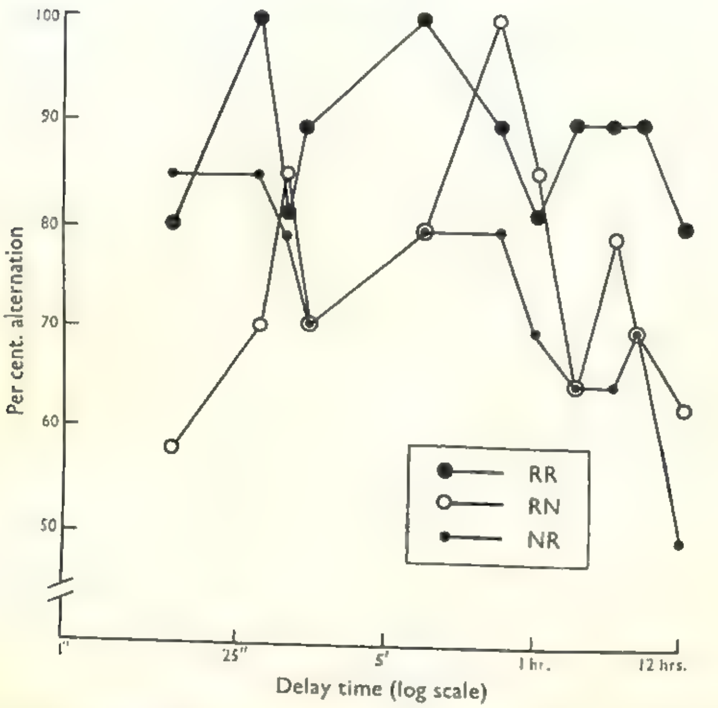
Percentage alternation as a function of delay is shown in Figure 3. Since the delays are not equally spaced, trend cannot be analysed by orthogonal polynomials.

TABLE II

INDEX OF DEGREE OF PREFERENCE SHOWN BY EACH RAT

Group RN		*Group RR*		*Group NR*	
Rat	*Index*	*Rat*	*Index*	*Rat*	*Index*
1	—	1	3	1	—
2	1·26	2	1	2	3·05
3	0	3	2	3	2·24
4	1	4	2·71	4	3·74
5	0	5	1	5	—
6	0·58			6	—
7	0·71			7	1

FIGURE 3



Alternation as a function of delay, for groups rewarded only on the first trial of a pair (RN), the second (NR) and both (RR).

Instead, linear trend coefficients are obtained by expressing the log delays as deviations from the mean log delay, and finding, for each rat, the sum of the products of each delay interval so expressed and the number of repetitions at that delay. This gives the regression coefficient of alternation on delay multiplied by a constant which is the same for all rats. The differences between groups in total alternation and trend are not significant by the Kruskal-Wallis test ($p > 0{\cdot}05$).

Alternation and stage of training. A comparison between frequency of alternation on days 51–61 and frequency on days 62–72 suggests that alternation has reached a stable level. The rats of group RN show 76·6 per cent. alternation during the first

block of 11 days, 74·0 per cent. during the second; group RR shows 87·3 per cent. and 89·1 per cent.; while group NR shows 72·7 per cent. alternation during both halves.

Analysis of repetition trials. Using the same index (I_{pref}) as before, the differences between RN, and both NR and RR are again significant.

Discussion

These results confirm the results of the experiments, referred to above, in which no decline in alternation with training was found. They also confirm Walker (1956) and Zeaman and House (1951) who found only a gradual decline in alternation with delay. But they conflict with the predictions from the theories considered in the Introduction, which suggests that these should be rejected as explanations of spontaneous alternation. Montgomery's (1952) theory may be saved in part by supposing that there is sufficient secondary reinforcement on the second trial of a pair for group RN to maintain alternation; but it still cannot account for alternation at long delays in highly trained rats.

It is possible that spontaneous alternation is a manifestation of an exploratory tendency, even though the details of Montgomery's account must be changed. For instance, the exploratory tendency may not decline even when the maze becomes familiar to the rat, either because some information is always lost over 24 hr. so that the maze is never fully explored, or for some other reason. Alternatively, it is possible that the exploratory tendency does decline, but that alternation continues as a learned habit which can be based upon secondary reinforcement. We need not suppose, as Montgomery did, that the learned habit involves a different kind of alternation (i.e. response instead of stimulus alternation), and there is therefore no reason to expect poorer alternation at longer delays.

The hypothesis that alternation only continues as a learned habit may be tested by comparing rats rewarded on all or some trials with rats not rewarded at all; the latter should show a greater decline in alternation with training.

Experiment II

Method

Subjects. Fifteen female hooded rats, about 5 months old at the beginning of the experiment, were used. They were bred from stock obtained from N.I.M.R., Mill Hill, London, and were frequently handled from birth. They were thus very tame, which was essential, since some of them were to run without reward. The rats had previously been used in an experiment involving five pairs of unrewarded trials in the T-maze used in this experiment.

Apparatus. The T-maze was that used in Experiment I; the only difference was that the lighting was provided by three 30-watt strip-lights, 3 in. above the Perspex roofs; the bulb holders were attached to the roofs.

Procedure. The rats were divided into three equal groups, equated on the basis of the amount of alternation in the previous experiment for which they had been subjects. The groups were randomly assigned to three conditions:

Group 100: Rewarded on both of the daily pair of trials (as for group RR of the last experiment).

Group 0: Never rewarded.

Group 50: Rewarded on 50 per cent. of the trials. They could be rewarded on both trials of a pair, the first only, the second only, or neither; each of these four conditions was given once over each block of four trials, and on any one day each condition was given to at least one rat in the group.

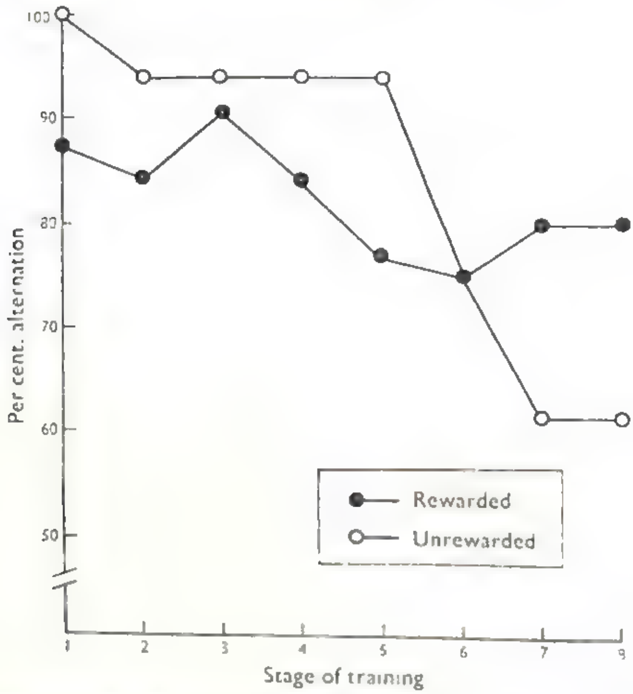
The experiment was continued for 32 days with one pair of trials per day, and 15 sec. between trials of a pair. Details were as for the previous experiment, except that the food holes contained 45 mg. sugar pellets, and the rats had food in the living cages for 1 hr., starting 15 min. after the day's session.

One rat from the 0 group died during the experiment and its results are not included.

Results

There were no significant differences, either in total alternation, or trend of alternation with training, between the two rewarded groups (50 and 100), and therefore these groups were combined to test the effect of reward as against no reward. Figure 4 shows alternation as a function of training. The rewarded and

FIGURE 4



Per cent. alternation over each block of 4 days, for rewarded and unrewarded rats.

unrewarded (0) groups do not differ significantly with respect to total alternation (one-tailed $p > 0{\cdot}05$ by the Mann-Whitney U-test). However, the rewarded rats show significantly less linear trend (decline in alternation) with training; comparing linear trend coefficients calculated by means of orthogonal polynomials, Mann-Whitney $U = 5\frac{1}{2}$, two-tailed $p < 0{\cdot}05$. All the rats in the 0 group show a negative trend, but since there are only four rats in this group the trend is not significant by the Wilcoxon T-test (one-tailed $p = 0{\cdot}0625$). Trends in the other two groups do not approach significance.

Analysis of repetition trials. As with groups RR and NR in the previous experiment, the rats tended to repeat according to a position preference; differences between the groups (using I_{pref}, defined above, as the index of position preference) do not approach significance.

Discussion

Since there is significantly more decline in alternation among the unrewarded rats, compared with the rewarded, this result is consistent with the views that alternation is initially part of the exploratory tendency, and only continues once the maze is explored if the rat is rewarded (primarily or secondarily) for alternating. Unlike the other theories considered, this theory does not predict any differences in

alternation as a function of reward schedule (provided the rat is rewarded), nor does it predict any interaction between reward schedule and delay under the conditions tested. Such differences were not found in the first experiment and we may therefore tentatively accept the theory as being the one best supported by the evidence.

There are two aspects of the results of the two experiments in this paper that have not so far been considered:

(i) Rats of group RN do not choose according to a position preference when they fail to alternate, while rats of all other groups do seem to do this. A possible explanation of this is as follows: Suppose a rat of group RN chooses *entirely* according to a position preference; then on the first trial it will be forced to choose the preferred arm on 50 per cent. of trials and be rewarded on all of them, while on the second trial, which is free, it will choose the preferred side on 100 per cent. of trials and will never be rewarded; it will thus be rewarded after 33 per cent. of such choices, when both forced and free trials are considered together. The non-preferred side, however, will only be chosen on forced trials, and will be rewarded on 100 per cent. of such choices altogether. With such a lack of balance in percentage reinforcement, giving a strong bias in favour of the non-preferred arm, the position preference should rapidly diminish, and this should happen to some extent even if the rat chose only sometimes according to a position preference, and according to the alternation tendency at other times. In the other groups the reward schedule would not favour the non-preferred arm, and therefore the same preference would continue to determine choice on trials on which the rat failed to alternate. Thus if it is assumed that rats start off with a preference for one arm or the other, all the rats, except those of group RN, should continue to manifest the same preference.

(ii) Rats of group RR in the first experiment show an increase in alternation with training compared with the other groups. This was not found, however, with group 100 in the second experiment, so that it cannot be considered a very reliable phenomenon; nor, however, does its occurrence in the first experiment seem likely to be due to chance, since similar findings have occasionally been reported (Walker, 1956; Montgomery, 1952). Thus it appears that under some conditions there will be an *increase* in alternation with training; but the data available are not adequate to determine why this should be so.

The author would like to thank the Medical Research Council for financial support; Mr. A. J. Watson, who supervised this research, for much helpful discussion and advice; and Professor O. L. Zangwill for his continued encouragement

REFERENCES

DEUTSCH, J. A (1960). *The Structural Basis of Behaviour*. London: Cambridge University Press.

FISHER, R. A. (1958). *Statistical Methods for Research Workers*. Edinburgh: Oliver & Boyd.

MONTGOMERY, K. C. (1951). Spontaneous alternation as a function of time between trials and amount of work. *J. exp. Psychol.*, **42**, 82–93.

MONTGOMERY, K. C. (1952). A test of two explanations of spontaneous alternation. *J. comp. physiol. Psychol.*, **45**, 287–93.

ROTHKOPF, E. Z., and ZEAMAN, D. (1952). Some stimulus controls of alternation behaviour. *J. Psychol.*, **34**, 235–55.

SIEGEL, S. (1956). *Non-parametric Statistics for the Behavioural Sciences*. New York: McGraw-Hill.

Still, A. W. (1966). Repetition and alternation in rats. *Quart. J. exp. Psychol.*, **18**, 103–8.

Still, A. W. (1967). The use of orthogonal polynomials with non-parametric statistics. To be published.

Sutherland, N. S. (1957). Spontaneous alternation and stimulus avoidance. *J. comp. physiol. Psychol.*, **50**, 358–62.

Walker, E. L. (1956). The duration and course of the reaction decrement and the influence of reward. *J. comp. physiol. Psychol.*, **49**, 167–76

Walker, E. L. (1958). Action decrement and its relation to learning. *Psychol. Rev.*, **65**, 129–42.

Weitz, J., and Wakeman, M. L. (1941). "Spontaneous" alternation and the conditioned response. *J. comp. Psychol.*, **32**, 551–62.

Wingfield, R. C., and Dennis, W. (1934). The dependence of the rats' choice of pathways upon the length of the daily trial series. *J. comp. Psychol.*, **18**, 135–47.

Zeaman, D., and House, B. J. (1951). The growth and decay of reactive inhibition as measured by alternation behaviour. *J. exp. Psychol.*, **41**, 177–86.

Manuscript received 9th October, 1965.

THE CAPACITY FOR GENERATING INFORMATION BY RANDOMIZATION

BY

A. D. BADDELEY

From the Medical Research Council Applied Psychology Research Unit, Cambridge

Response selection was studied independently of the stimulus by asking subjects to generate random sequences of letters or numbers. Experiment 1 varied rate of letter generation from ½ sec. to 4 sec. per item and showed that the redundancy of the sequence increased linearly with rate. Experiment 2 added random generation of letters as a secondary task to paced card sorting. Information load per card was varied from 1 through 2 to 4 to 8 alternatives, with sorting rate held constant. As predicted, the redundancy of the sequences generated increased linearly with sorting load. Experiment 3 varied number of items to be randomized. Rate of random generation increased systematically from 2 to 4 to 8 alternatives, but levelled out beyond this point, showing no difference between 16 and 26. In general, these results suggest a response-selection mechanism of limited informational capacity.

There is considerable evidence from experiments on decision-taking, guessing and gambling that man seldom produces a random sequence of events (Tune, 1964). There is further evidence that consistent biases occur even when subjects are specifically instructed to be random.

Baddeley (1962) showed that similar biases occurred when subjects were asked to call out a random sequence of digits or letters at a paced rate of one item per sec. However, when asked to select the random digit sequences from a mixture containing both sequences generated by other subjects and sequences based on random number tables, subjects selected those based on the random number tables, suggesting that their *concept* of randomness was not at fault. Furthermore, when allowed to produce a sequence at their own preferred rate, subjects went more slowly and produced sequences that were considerably less biased. This latter result suggested that the bias might be due to limitations in the capacity of the subject's response-selection process. Such a limitation has been suggested by attempts to apply information theory to performance in reaction time studies (Leonard, 1960; Welford, 1960).

If the biases shown when subjects try to produce a random sequence of events under a time stress are due to the limited informational capacity of the under-lying response-selection mechanism, then the randomness of such a sequence should be systematically related to (*a*) the rate of generation, and (*b*) the number of alternatives which the subject is required to randomize. Experiments 1, 2 and 3 explore these predictions.

Experiment I

Method

The task required the subject to call out a sequence of 100 letters of the alphabet paced by a timer clicking at each of four rates; one letter per ½, 1, 2 and 4 sec. The subjects were told to imagine they were drawing letters from a hat one at a time, calling them out, and replacing them, so that on each draw any of the 26 letters of the alphabet was equally likely to be selected. It was further pointed out that such a sequence would be completely jumbled and would not therefore be likely to comprise English words or alphabetic sequences such as ABC or XYZ. A practice run of about 20 responses (each of one letter), was then given at the 1-sec. rate to ensure that the subject had understood the instructions and could keep pace with the timer. This was followed by the main test

which involved one sequence of 100 responses at each of the four rates, a 5-min. break, then another 100 responses at each rate. The order of presentation of conditions was randomized for each subject with the constraint that no subject began with the ½-sec. rate, which was the only condition in which subjects had any difficulty in keeping pace. The letters were written down by the experimenter as they were spoken in all conditions except the fastest, in which they were tape recorded and transcribed later. Twelve males aged between 20 and 30 served as subjects.

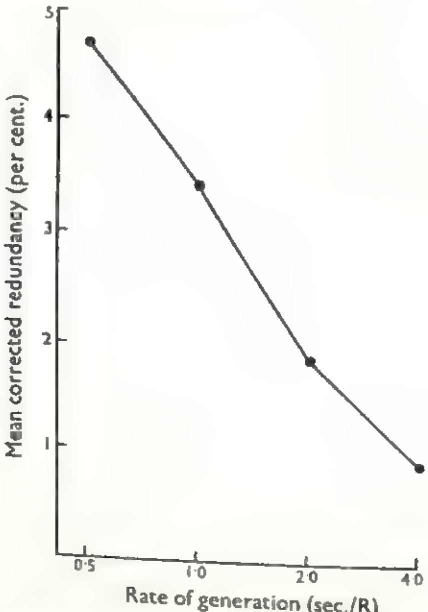
Results

Randomness was estimated in terms of both single letters and digrams or pairs. An information measure was computed, based on the relative frequency with which individual letters were used. A sequence of letters contains the maximum first-order information when each letter is used with equal frequency; as the frequency with which individual letters are used becomes unbalanced, the sequence becomes more redundant. An information score was computed using the formula:

$$H = log_2 n - \frac{1}{n} (\Sigma n_i \, log_2 n_i)$$

where H is the information per letter in bits, n is the number of responses in the sample and n_i ($i = 1, 2 \ldots\ldots 26$) is the frequency with which the ith letter of the alphabet was used (Attneave, 1959, p. 24). Since it is advisable to have at least six observations per cell (Miller, 1955), and 26 cells are involved, the subjects' two runs of 100 responses were combined at each level. Even so, a bias would be expected with a sample of this size, and the values obtained were therefore corrected using the Miller-Madow correction (Miller, 1955). This information score was then converted into per cent. redundancy (Attneave, 1959), and is shown in Figure 1.

FIGURE 1



First-order redundancy as a function of rate of generation.

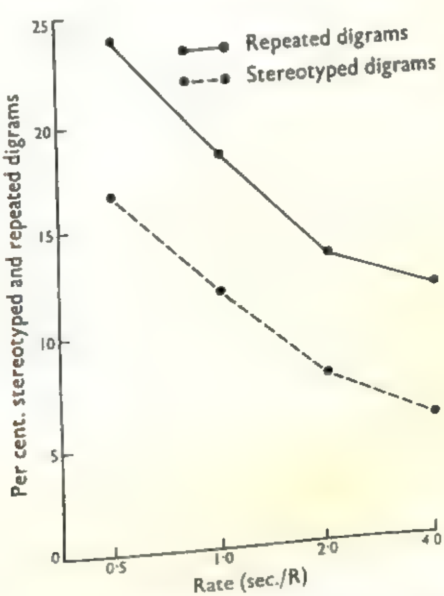
It is clear that as rate increases, the frequency with which individual letters occur becomes more unbalanced, and the sequence more redundant. The linear component is large (98·5 per cent.), and significant ($F = 134{\cdot}4$, $d.f. = 1, 2$. $p < 0{\cdot}01$). A

Freidman test indicated significant differences between the four rates ($p < 0{\cdot}001$), and comparisons between the various rates using the Wilcoxon test showed all the differences between conditions to be significant at beyond the 0·05 level.

The next measure attempted to assess the amount of digram information in each letter sequence. Ideally, the information measure used for single letters should be applied to the frequency with which each possible digram or pair of letters occurred in the sample. However, with 676 possible digrams, a reasonably reliable estimate would require more than 4,000 responses. An approximation to this measure was therefore used, which involved simply counting the number of repeated digrams produced in generating 100 responses. This was scored using a 26×26 matrix in which each cell represents one of the 676 possible digrams. If the subject were "truly" random, he would be expected to repeat about seven digrams in generating 100 letters. The more he repeats himself, the less random he is.

A second digram measure was based on the frequency with which the subject produced digrams which follow the alphabetic stereotype (e.g. AB, KL, or YZ). Both these digram measures are crude and are open to the objection that the proportion of digram constraint that is attributable to first-order redundancy cannot be assessed. If no digram score is used, however, the subject could adopt a strategy such as reciting the alphabet from A to Z, which, though stereotyped would give him a perfect first-order information score.

FIGURE 2



Number of stereotyped and repeated digrams as a function of rate of generation.

The two digram scores were computed for each run of 100 responses and are shown in Figure 2. In the case of the digram repetition score, the linear component accounts for 95·2 per cent. of the variance ($F = 40{\cdot}0$, $d.f. = 1, 2$. $p < 0{\cdot}05$). The linearity is less convincing than in Figure 1, particularly at the slowest rate. This is probably due to a "ceiling" effect as some subjects were already approaching optimal performance at the 2-sec. rate. A Freidman test indicated overall differences

between the four rates ($p < 0.001$). Comparing the rates using a Wilcoxon test, all differences between conditions again proved significant at beyond the 0·05 level. The score based on the percentage of stereotyped responses has a linear component which accounts for 97 per cent. of the variance ($F = 67.6$, $d.f. = 1, 2$. $p < 0.05$). A Friedman test again showed significant differences between the four rates ($p < 0.001$) and comparisons using the Wilcoxon test showed all conditions to be significantly different from each other at beyond the 0·05 probability level.

On all three measures of randomness there is a clear relationship between rate of generation and amount of information per letter. When plotted on a binary log scale, this relationship approximates a straight line, though both the digram scores show a tendency for performance to level off at the 4-sec. rate. In both these cases, however, a perfect straight line relationship would have required several subjects to produce even fewer stereotyped responses and repeated digrams than would be found in the sort of random sequence he was instructed to produce. A levelling off was therefore almost inevitable. In general, however, these results suggest that the rate of information generation is constant across conditions, the subject can only increase his rate of responding by reducing the amount of information per response. This supports the hypothesis that the response-selection mechanism underlying random generation has a limited informational capacity; the less probable the response the longer it will take to select.

Experiment 2

Experiment 1 examined the limited capacity hypothesis by varying the load imposed by the generating task itself. Experiment 2 approaches the problem by using random generation as a measure of the load imposed by a second task. The task used involved sorting cards into 1, 2, 4 or 8 categories at a constant paced rate. Crossman (1953) has shown that the load imposed by this task increases as a linear function of log number of categories. If both random generation and card sorting employ the same information channel, and if this channel is of limited capacity, then the redundancy of the letter sequences generated should be a function of number of sorting alternatives. With sorting time constant, there should be a linear relationship between the amount of information in the sorting task and the redundancy of the letter sequence produced.

Method

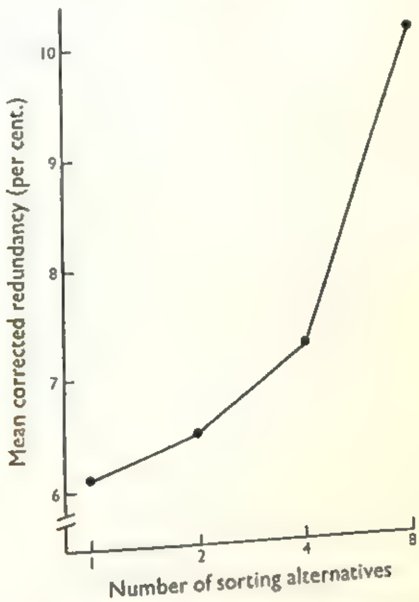
Subjects sorted packs of 100 playing cards into the following categories: 2-choice, kings and eights; 4-choice aces, twos, fours and eights; 8-choice, aces, twos, threes, fours, fives, sixes, eights and nines. The 1-choice situation employed one of the multi-choice packs with the subject instructed to ignore card values. The cards were held face down and sorted with the preferred hand into a sorting frame. This consisted of ten compartments, each 6 in. wide, 12 in. long, 6 in. deep and open at the end facing the subject. The middle two, four, or eight compartments were used, and when in use each compartment contained an appropriate marker card. In each case the arrangement was systematic with the smallest value on the left and the highest value on the right.

All subjects sorted in time with a metronome ticking at a 2-sec. rate and called out a letter with each sorting response. Instructions for randomizing were as in Experiment 1. All subjects were tested on two successive days. On Day 1 they were given a short practice run on generating alone, they were then required to generate two sequences of 100 responses at each level of sorting difficulty. On Day 2 they produced a further two sequences of 100 responses at each level of difficulty. Subjects were 12 males aged between 30 and 50; each subject performed the four conditions in a different random order.

Results

All subjects kept pace with the metronome and made a negligible number of errors. Any difference between the four conditions would therefore be expected to be shown in the generating task. This was analysed using the three scores described in Experiment 1.

FIGURE 3

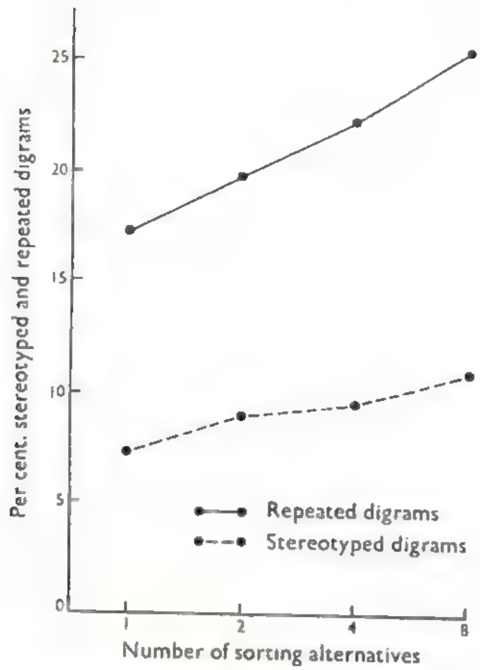


Letter generation while card sorting. First-order redundancy of letter sequence as a function of number of sorting alternatives.

The redundancy due to unequal frequency of using individual letters is shown in Figure 3, corrected for chance (Miller, 1955). A Friedman test showed overall differences among conditions ($p < 0{\cdot}01$), and the Wilcoxon test showed significant differences between the 1-choice condition and both the 4- and the 8-choice ($p < 0{\cdot}01$), between the 2-choice and 8-choice ($p < 0{\cdot}01$), and between the 4-choice and 8-choice conditions ($p < 0{\cdot}05$). Although the linear component of the relationship between randomness and number of sorting alternatives accounts for 86 per cent. of the variance, it is not significant ($F = 12{\cdot}3$, $d.f. = 1, 2$. $p > 0{\cdot}05$), and the relationship appears to be positively accelerated. The reliability of this tendency was examined by taking the increase in redundancy between the 1- and 2-alternative conditions, and comparing it with the equivalent increase between the 4- and 8-alternative conditions. A Wilcoxon test indicated that this difference was not reliable ($T = 17$, $N = 12$, $p > 0{\cdot}05$), suggesting that the deviation from linearity is not significant.

Figure 4 shows the influence of number of sorting alternatives on digram information, measured in terms of number of repeated digrams per 100 responses. Again the Freidman test showed significant differences between conditions ($p < 0{\cdot}01$). The Wilcoxon test showed all the differences between conditions to be significant at beyond the 0·05 probability level. The relationship between digram information and number of sorting alternatives again has a linear component which is large (98·8 per cent.), and significant ($F = 164{\cdot}5$, $d.f. = 1, 2$. $p < 0{\cdot}01$). The percentage of stereotyped responses produced in the four conditions is also shown in

FIGURE 4



Letter generation while card sorting. Number of stereotyped and repeated digrams as a function of number of sorting alternatives.

Figure 4. A Freidman test again indicated significant differences among conditions ($p < 0{\cdot}01$), and the Wilcoxon test showed a significant difference between the 1-choice condition and each of the others ($p < 0{\cdot}01$), and between the 2- and 8-choice conditions ($p < 0{\cdot}05$). The linear component of the relationship accounts for 96·0 per cent. of the variance ($F = 47{\cdot}5$, $d.f. = 1, 2$. $p < 0{\cdot}05$).

The results of both this and the previous experiment support the hypothesis that random generation depends on a response-selection mechanism of limited informational capacity. The less time or spare capacity available, the less random the output.

Experiment 3

This investigates the limited-capacity hypothesis further by studying the influence of number of alternatives on rate of randomizing. As number of alternatives is increased, the information per random response should increase, and with it the time required to select the response.

Method

The experimental design used is shown in Table 1. In each of four conditions subjects were placed in one of five groups, each group being assigned a different number of alternatives (2, 4, 8, 16 or 26). The assignment of subjects was initially random, but was subsequently biased by assigning more subjects to groups given eight or more alternatives when it became obvious that differences between these groups were much less clear-cut than differences between groups with smaller numbers of alternatives.

A group testing procedure was used in which the subject was required to write his responses on a sheet containing five columns of 32 cells each. All subjects had three runs of 2 min. and used a separate response sheet for each. "Paced" subjects were asked to try to write in time to a metronome beating at a 1-sec. rate, while unpaced subjects

TABLE I

Design of Experiment 3, Showing for Each Condition the Items Randomized and Number of Subjects Tested

		Number of alternatives					Total
		2	4	8	16	26	
Unpaced	I Letters	A & B N = 20	A — D N = 19	A — H N = 28	A — P N = 28	A — Z N = 25	120
	II Numbers	29 & 30 N = 18	28 — 31 N = 18	26 — 31 N = 19	22 — 37 N = 19	17 — 42 N = 18	92
"Paced"	III Letters	A & B N = 18	A — D N = 17	A — H N = 29	A — P N = 33	A — Z N = 27	124
	IV Numbers	29 & 30 N = 19	28 — 31 N = 21	26 — 33 N = 22	22 — 37 N = 20	17 — 42 N = 19	101

were simply asked to work as quickly as possible without becoming non-random. In two conditions subjects were asked to randomize letters, and since the smaller categories might be expected to have a greater proportion of high-frequency letters, two further conditions were run in which subjects randomized numbers. In order to avoid a mixture of one and two-figure numbers, 29 and 30 were selected for the two alternative case, and larger vocabularies were produced by expanding the series in both directions (see Table I). The subjects, 437 young males, were tested in groups of from 10 to 30 with subjects given different numbers of alternatives tested simultaneously.

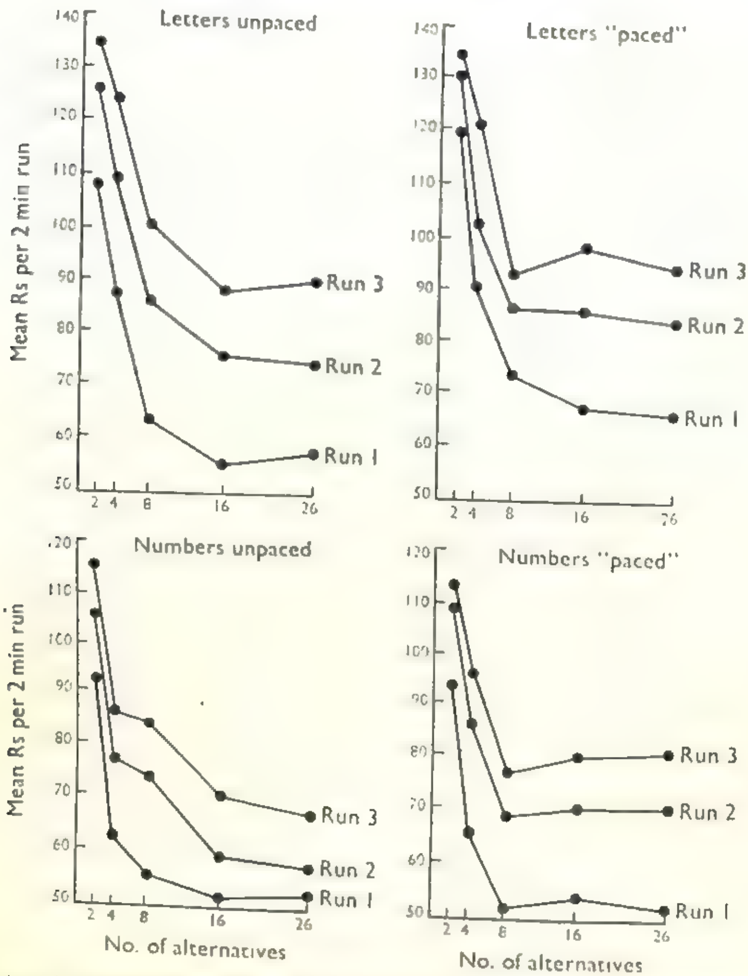
Results

Rate of generation. Figure 5 shows the rate at which subjects generated in the four groups. In the "paced" conditions subjects appeared to be unable to keep in time with the metronome, and in both "paced" and unpaced conditions the rate of responding was clearly dependent on number of alternatives. Scores for the three runs were combined and comparisons made using the Mann-Whitney test. They showed: (*a*) A significant difference between the 2- and 4-alternative groups for all four conditions ($p < 0{\cdot}02$ in each case). (*b*) A significant difference between the 4- and 8-alternative groups for all except the unpaced number condition ($p < 0{\cdot}02$). (*c*) A difference between the 8- and 16-alternative groups for both of the unpaced conditions ($p < 0{\cdot}05$ one-tail). (*d*) There was no significant difference in any condition between the 16- and 26-alternative groups. There was a consistent increase in rate of responding with practice in all conditions. This was shown by all but 27 of the 437 subjects tested, and was thus highly significant. A Wilcoxon test between equivalent "paced" and unpaced groups showed significantly faster performance with pacing ($p < 0{\cdot}05$). A similar analysis showed consistently faster performance with letters than with digits ($p < 0{\cdot}01$), probably because the numbers involved writing two characters per response, whereas letters involved only one.

The rate at which subjects generate both numbers and letters is clearly dependent on the number of alternatives they are asked to use. From 2 to 8 alternatives the relationship between rate and size of vocabulary approximates reasonably closely to a straight line when plotted on a log scale, for all four conditions, and in the unpaced conditions the relationship is reasonably linear up to 16 alternatives. The relationship clearly breaks down at some point between 8 and 16, with subjects

randomizing 26 alternatives just as rapidly as 16. The exact point of breakdown is not clear. In the "paced" conditions it appears to occur at 8, in the unpaced conditions somewhat later, although even in these conditions, the difference between 8 and 16 alternatives is barely significant.

FIGURE 5



Rate of random generation as a function of number of alternatives.

Randomness. Efforts to compare the randomness of sequences using different numbers of alternatives did not prove very fruitful. Neither the stereotype score nor the digram count used previously is suitable for comparing conditions when number of alternatives is not held constant. An informational measure would be suitable given a large enough sample. Clearly a digram analysis which would require up to 4,000 responses per subject in some conditions is impracticable with the present data. An analysis based on single letter frequency seemed more feasible and was tried on one condition, "paced" letters. Well over half the subjects failed to show any redundancy. There was a tendency for redundancy to increase with number of alternatives, but even with 26 alternatives, mean redundancy was only slightly over 2 per cent. In view of the small and variable sample size, especially

where large numbers of alternatives were used, the validity of such an analysis seems very dubious (Miller, 1955). A more profitable approach would seem to be to collect much larger samples of responses under the more strictly controlled conditions of individual testing.

Discussion

The three experiments just described all suggest that subjects find difficulty in producing random sequences of letters, and that their deviation from randomness occurs in a very lawful manner. Such deviations have previously been noted, and have been variously ascribed to the subject's concept of randomness (Chapanis, 1953; Dale, 1960; Richard, 1962), the limitations of immediate memory (Tune, 1964), and to the subject's level of arousal (Weiss, 1964).

One of the limits of the subject's performance is clearly set by his concept of randomness, and in some situations this is certainly an important factor. It seems highly unlikely, however, that he should vary his *concept* of randomness with the rate at which he is required to randomize; indeed most of the present subjects were acutely aware of their departures from randomness, but apparently could not avoid them without slowing down their rate of generation. This awareness of non-randomness also makes Tune's immediate memory explanation rather unconvincing. Tune argues that "the essential parts of making a random series of selections are the ability to use a knowledge of past responses and remembering them" (Tune, 1964, p. 708). This seems debatable for two reasons; first, because a random sequence is one in which the events are independent; it is precisely because the subject has a memory that on this task he is inferior to a random number table. Secondly, if non-randomness were due to the subject's forgetting his previous response, performance should deteriorate as the time between responses increases. Experiment 1 shows the opposite to be the case.

A third explanation of biases in random generation is suggested by Weiss (1964), who used only two alternatives and compared the performance of schizophrenics with normals. He found deviation from randomness to be particularly marked in the schizophrenics, an observation that was also made by Yavuz (1963) who used alphabet randomization. Weiss attributes this non-randomness to the subject's failure to pay attention. He attributes this to the low level of arousal assumed to accompany random generation, which he describes as an unstimulating task. This explanation is not compatible with the results of any of the present experiments; an increase in rate, number of alternatives or in the difficulty of a supplementary task would all be expected to increase arousal, and thus on Weiss's hypothesis should improve performance. In each case the opposite result occurs.

The present results, then, suggest that neither the subject's concept of randomness, his immediate memory span, nor his level of arousal provide an adequate explanation of random generating behaviour. They do, however, suggest that the subject behaves as if he has limited informational output capacity. What does this imply? The most conservative conclusion is that, given this narrowly specified situation, subjects appear to behave in a systematic manner. However, it seems unlikely that they should have a special mechanism or even a common strategy for a task as improbable and unfamiliar as random generation. It seems more parsimonious to assume that the selection of a response in this situation involves the same basic process as occurs in selecting a response in other situations such as recalling a name or performing a skilled motor response. What probably differs in a recall situation is the subject's criterion for accepting or rejecting a response. In learning, where the subject is trying to reduce randomness a response will probably be accepted if

it has previously occurred in that context (*cf.* Miller and Frick, 1949). In a randomizing situation, the subject should reject familiar responses unless he has insufficient time to select an alternative. However, apart from this difference in criterion it seems not unreasonable to suppose that producing a response involves a similar process in both learning and random generating. How well then, do the data fit this hypothesis?

Experiments 1 and 2 produce results that are consistent both with a simple informational model and with analogous reaction time studies. Experiment 3 shows a discontinuity in the relationship between rate of generation and number of alternatives which is clearly not consistent with such a simple model. Whether or not it is consistent with the equivalent reaction time situation is, however, less clear. Relatively few reaction time studies have used more than 10 alternatives. Of those doing so, several have used highly compatible S–R pairings, producing a slope which is so slight that a discontinuity would be very difficult to detect (Davis, Moray and Treisman, 1961; Pierce and Karlin, 1957). Of the remainder, two studies show a linear relationship between reaction time and number of alternatives extending beyond 10, a card sorting study by Crossman (1953), and an experiment on rate of reading nonsense syllables by Conrad (1962). Both of these situations, however, involved relatively unpracticed S–R combinations. Level of practice may be an important factor, since Schmidtke (1961) claims to have found a simple linear relationship between reaction time and number of alternatives during the early stages of learning, which changed with practice to give a discontinuous relationship, with reaction time increasing up to 10 alternatives, and from then on being roughly constant. A similar conclusion is reached by Seibel (1963, p. 222) who, in trying to fit together data from several separate key-pressing experiments suggests "a function which increases from 1 to approximately 3 bits (2 to 8 alternatives), and shows little further increase from 3 to almost 10 bits (8 to almost 1,024 alternatives)." The data therefore are still equivocal. However, if, as has been suggested, random generation and choice-reaction share a common response selection mechanism, future studies should show a clear discontinuity in the relationship between reaction time and number of alternatives, and demand a rather more complex informational model than is usually proposed. Should this prove to be the case, random generation seems likely to provide a more effective method of studying the response-selection process since by eliminating the stimulus, it avoids the difficult problem of S–R compatibility which makes the accurate prediction of reaction-time results so difficult. Whether or not this attempt to predict the results of reaction-time experiments on the basis of the relationship observed in random generation proves fruitful, however, random generation itself appears to be described most simply in terms of a response-selection process of limited capacity.

The author wishes to thank Dr. R. Conrad for his valuable comments and constructive criticism.

References

Attneave, F. (1959). *Applications of Information Theory to Psychology*. New York: Holt.

Baddeley, A. D. (1962). Some factors influencing the generation of random letter sequences. *Med. Res. Council. Appl. Psychol. Res. Unit Rep.*, No. 422/62.

Chapanis, A. (1953). Random number guessing behavior. *Amer. Psychologist*, **8**, 332.

Crossman, E. R. F. W. (1953). Entropy and choice times: the effect of frequency unbalance on choice responses. *Quart. J. exp. Psychol.*, **5**, 41–51.

Conrad, R. (1962). Practice, familiarity and reading rate for words and nonsense syllables. *Quart. J. exp. Psychol.*, **14**, 71–6.

DALE, H. C. A. (1960). A study of subjective probability. *Brit. J. stat. Psychol.*, **13**, 19–29.

DAVIS, R., MORAY, N., and TRIESMAN, A. (1961). Imitative responses and the rate of gain of information. *Quart. J. exp. Psychol.*, **13**, 78–89.

LEONARD, J. A. (1960). Choice reaction time experiments and information theory. *Fourth London Symposium on Information Theory*. London: Butterworth.

MILLER, G. A. (1955). Note on the bias of information estimates. In QUASTLER, H. (Ed.), *Information Theory in Psychology*. Glencoe, Illinois: Free Press.

MILLER, G. A., and FRICK, F. C. (1949). Statistical behavioristics and sequences of responses. *Psychol. Rev.*, **56**, 311–24.

PIERCE, J. F., and KARLIN, J. E. (1957). Reading rates and the information rate of a human channel. *Bell System tech. J.*, **36**, 497–516.

RICHARD, J. F. (1962). Étude de la stereotypie des responses dans les compartements au hasard en fonction de l'age. *Bull. Psychol. Gr. Étud. sci.*, **16**, 285–96.

SCHMIDTKE, H. (1961). Zur Frage der informations theorischer Analyse von Wohlreations-experimenter. *Psychol. Forsch.*, **26**, 157–78.

SEIBEL, R. (1963). Discrimination reaction time for a 1,023-alternative task. *J. exp. Psychol.*, **66**, 215–26.

TUNE, G. S. (1964). A brief survey of variables that influence random-generation. *Percep. mot. Skills*, **18**, 705–10.

WEISS, R. L. (1964). On producing random responses. *Psychol. Rep.*, **14**, 931–41.

WELFORD, A. T. (1960). The measurement of sensory-motor performance: survey and reappraisal of twelve years' progress. *Ergonomics*, **3**, 189–230.

YAVUZ, H. S. (1963). The production of random letter sequences in schizophrenics. *J. Psychol.*, **56**, 171–3.

Manuscript received 20th October, 1965.

ON THE HANDLING OF HEAVY BIAS IN A SELF-PACED TASK

BY

J. A. LEONARD,* R. C. NEWMAN and A. CARPENTER

From the Medical Research Council Applied Psychology Research Unit, Cambridge

Following an earlier observation that systematic variations in performance on a task with heavy frequency imbalance was correlated with fluctuations in the amount of bias in the input programme, an experiment was carried out to establish the roles of long and short term sampling. Two groups of subjects were trained on a self-paced, five-choice task. One group's input had an average of 68 per cent. bias on one source, the other had 44 per cent. on the same source. Analysis of data was carried out on three levels of local bias for each condition, one level being identical for both conditions. It was found that responses to the biased stimuli were determined by the average bias in each input sequence and not by moment to moment variations in that bias. The effect observed originally can therefore be accounted for in terms of a relatively simple additive model which includes the "repetition effect" first described by Bertelson.

Introduction

For some time now, we have been carrying out a series of experiments on the acquisition and maintenance of speed and accuracy in a keyboard task (Leonard and Newman, 1965). A self-paced serial choice apparatus developed from one described in Leonard (1959) was used: one signal appeared with a probability of 0·68, and four others with a probability of 0·08 each. A total stimulus sequence of 10,000 stimuli was broken up into units consisting of blocks of 5×50 stimuli. Since the generating procedure was completely random, this meant that some blocks would have more and others less bias than the average specified. We soon noticed, and subsequently confirmed, that there was a negative correlation between the amount of bias in, and the time required to complete a block of trials.

The purpose of the present paper is to show by a more detailed examination of response times just how this effect was achieved, and to enable us to make a more general statement about similar situations.

What has to be explained? In all but one of the conditions of our experiments, the subjects could only see one stimulus at a time and did not know whether they were about to be given a block of trials with heavy, medium, or light "local" bias. Yet when there were more than an average number of biased stimuli present, subjects took less total time per block than when there were fewer than the average. There would appear to be two possible kinds of explanation.

(1) The subjects learned to respond with a constant low reaction time to the biased signal, in which case the effect observed can be explained by the simple fact that blocks with a larger number of biased stimuli will contain a higher number of fast responses contributing to the total time per block than the blocks with a smaller number of biased signals. The pattern of responding to be found under those conditions would not be affected by moment to moment fluctuations in bias and we would ascribe it to "long-term sampling."

(2) Subjects might be able to adjust their performance on the basis of the immediate past, in which case we would expect to find a pattern of responding which changed with the amount of local bias present. This we would ascribe to "short-term sampling."

* Now at Psychology Department, University of Nottingham.

The case for the latter type of explanation is rather strong in view of the "repetition-effect" reported by Bertelson (1961) and its more formal exposition by Falmagne (1965). For one effect of introducing any frequency imbalance in a sequence of stimuli is to increase the frequency of "runs" of the biased stimulus, i.e. repetitions of the same stimulus. In a sense, it is the response to such runs which is at the heart of the matter under discussion.

Hyman (1953) and Crossman (1953) had both shown that with the number of alternatives kept constant the introduction of frequency imbalance speeded up performance relative to the equiprobable frequency situation. Hyman drew attention to the fact that whenever a stimulus had been repeated without the intervention of another stimulus, an unusually fast reaction time had been observed. Bertelson (1961), using an equiprobable two-choice task, varied the sequential structure in such a way that in one series there was a preponderance of "repetitions" while in another sequence there was a preponderance of alternations; the information conveyed was the same in both sequences. He found that subjects who had a lot of repetitions were faster than those who had a lot of alternations, and this he ascribed to facilitation due to the "repetition effect." A more detailed analysis of his data confirmed Hyman's observation and indicated a progressive decrease in reaction time as a function of position of stimulus within a "run." Falmagne (1965) developed a mathematical model of the linear-operator type which predicts an exponential decrease in reaction time as a function of the number of stimuli intervening before the same stimulus is repeated.

As far as choice reaction time data are concerned, almost all the reported effects of varying information content of inputs on choice reaction time could be accounted for by such a model and not by an information theory type of model. The distinction lies in the fact that Falmagne's model explains performance in terms of the moment to moment experience of the subject while an information theory type of explanation deals with long-term stable expectancies set up prior to the time at which data are examined.

Finally, Laming (personal communication, 1965) carried out a series of experiments using inputs with frequency imbalance, and applying a combination of multiple regression and autocorrelation analysis to his data, concluded that performance at any given moment could be accounted for in terms of the immediate past experience of the subject.

Thus the evidence tends to favour an explanation in terms of subjects' ability to adjust performance in the light of immediate past experience. Just where one places the dividing line between immediate past experience and long-term experience is a moot point, but the crux of the matter is the difference between a pattern of responding which is relatively unaffected by moment to moment changes in the input, and a pattern which is affected by such changes.

For the experiment to be described, we decided to compare performance of two groups: one was to be trained on 68 per cent. bias and the other on a bias of 44 per cent. so that we could compare directly performance on trials common to both groups but which would be relatively low in local bias as far as the 68 per cent. group was concerned, and relatively high as far as the other group was concerned.

The input programme for this other group had to fulfil three criteria:

(1) The distribution of local bias in trials had to be sufficiently close to that of the 68 per cent. distribution to give us an adequate number of trials in the overlap area.

(2) The bias had to be sufficiently different from 68 per cent. for an overall difference in performance to be reasonably expected.

(3) The average bias had to be still sufficiently marked for subjects to treat the sequence as a biased one.

It will be seen that a programme with average bias of 44 per cent. met our requirements. The results presented in this paper will be pertinent to the following questions:

(i) how is the correlation between local bias fluctuations and performance achieved?

(ii) is performance on the "identical trials" determined by short-term sampling (in which case we would expect to find identical measures of performance on these trials for the two groups), or is it determined by long-term sampling (in which case we would expect to find reliable differences in performance measures)?

Method

Apparatus

A modified form of a self-paced serial choice-reaction apparatus described by Leonard (1959) was used. It had the following properties:—

(1) Stimulus sequences were programmed on a perforated tape.

(2) Both correct and incorrect responses advanced the programming device, i.e. a non-correction procedure was adopted.

(3) Errors were cumulated on a single counter.

(4) Time per 50 responses (i.e. one trial), was measured on a Standard Electric Timer to the nearest 0·01 sec., i.e. the timer ran continuously from the moment the experimenter pressed the start button to the moment the subject made the 50th response.

(5) Having pressed a key, the subject had to release that key before being able to press the same or another key.

(6) Pressing more than one key together scored a single error.

(7) The next stimulus was presented within rather less than 100 millisec. of a key being pressed. If the same signal was repeated, there was a brief blink on the display to achieve this effect.

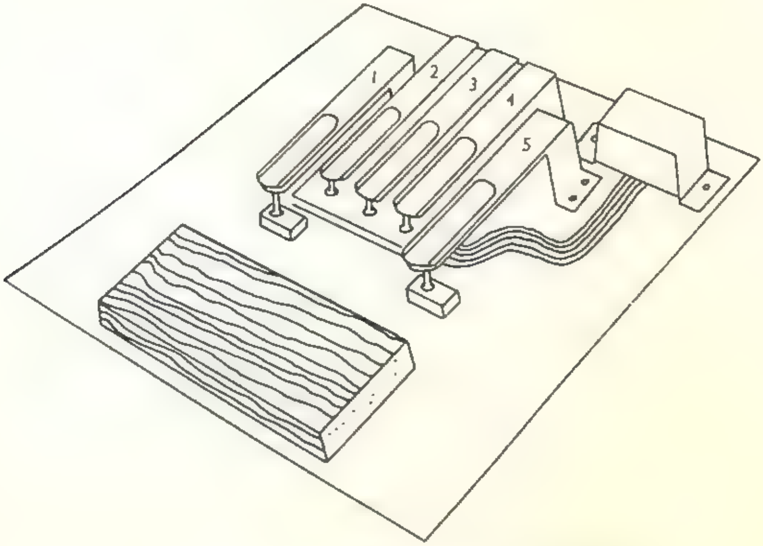
(8) The maximum rate of operation of the whole equipment was 10 items/sec.

(9) In addition to the overall time and error scores detailed records for each response were obtained with the aid of SETAR (Welford, 1952). The method adopted provided a single readout for each response giving the time elapsed since the last response to the nearest 0·01 sec.; a stimulus and a response code.

Display-control arrangements

The display consisted of a One Plane Digital Display Unit, Model 100 with 6·3 v. lamps, manufactured by Industrial Electronic Engineers (the supply circuit ensured that any bulb would be lit well within 100 millisec. regardless of the length for which it had previously been off (Carpenter and Leonard, 1960). With this unit any digit from 1–5 could be projected from the rear on to a $1\frac{1}{2}$ in. × 2 in. screen so that the digits, 1 in. high, all appeared in approximately the same place on the screen. The display was mounted at about the subjects' eye level in the frontal-parallel plane, and about 18 in. behind the keyboard. The five-finger keyboard used is shown in Figure 1. The centres of the keys were 1 in. apart, a pressure of about 4 oz. and a movement of about $\frac{1}{16}$ in. at the free end of the keys was required to operate the microswitch mounted beneath each key. A wooden block, $\frac{3}{4}$ in. high, $7\frac{1}{2}$ in. long and $1\frac{1}{2}$ in. wide served as a rest for the heel of the hand. This keyboard was used for five-finger operations, i.e. one finger to each key.

FIGURE 1



The keyboard used in the present experiment.

The S–R relationship was such that digit 1 corresponded to the key on the left of the board, digit 2 to the key next to it, and so on. Labels for the appropriate digits were stuck near the hinge of each key.

The subject sat in a noise attenuating cubicle which was ventilated by a fan. The experimenter sat outside the cubicle and could observe the subject through a one-way screen. The experimenter could talk to the subject via a loudspeaker.

Stimulus sequences

68 *per cent. bias on signal three.* This sequence was generated from Fisher's tables of random numbers so that on average signal three would appear 68 per cent. while each of the others would appear 8 per cent. of the time. A total of 24 blocks were produced, of these, 12 were used both backwards and forwards, two once forwards and twice backwards, and 10 forward only, making a total of 50 blocks. (No blocks were repeated within the second week's programme.)

44 *per cent. bias on signal three.* This sequence was generated by a programme from EDSAC so that, on average, signal three would appear 44 per cent., and each of the others 14 per cent. of the total time. The whole sequence of 10,000 stimuli thus generated was used with this exception: seven trials from the second half of the 68 per cent. programme were transplanted to appear in exactly the same ordinal position in the 44 per cent. programme. This meant that seven trials of the 44 per cent. programme were thrown out and some minor adjustments had to be made to accommodate the seven transplanted trials. The throw-outs were as nearly similar as possible to the transplants as far as incidence of stimulus three was concerned. The overall distribution of bias for 44 per cent. was thus hardly affected.

Procedure

Each subject attended for two 15 min. sessions on each of 10 days, once in the morning and once in the afternoon.

In each session subjects were first given 25 responses of five-choice equi-probable material. This was intended as limbering up and subjects were told about errors, if any, but not given their time score. They were then given their first block of five trials of 50 responses each. After each trial they were then given their total time and error score. In the 1½ min. intervals between first and second blocks the experimenter talked to the subject in the cubicle about his progress.

Instructions

Subjects were instructed to respond as fast as possible while not making more than one error per trial of 50 responses. Subjects were informed about the amount of bias to be expected and were warned of the difficulties of responding quickly to a biased signal while still maintaining accuracy.

Subjects

There were 11 subjects, six performing on the 44 per cent. programme, and five on the 68 per cent. programme. The subjects were all volunteers from the Royal Navy.

Treatment of data

(1) For the detailed analysis, only the second week's data was used.

(2) Mean times per response for successive blocks of trials and mean error percentages were worked out from the data sheets containing the raw data from timer and counter.

(3) The SETAR data obtained for the second week of the experiment were processed by EDSAC at the Cambridge University's Mathematical Laboratory. For each subject an output was obtained which summarized performance in three categories: High, Medium and Low Bias trials. The observed bias for each category is shown as a percentage in Table III. Within each category, the output gave the number of stimuli, the mean time and the standard deviation per response to each of the five stimuli, as well as a run analysis for stimulus three. This run analysis showed the number of entries in each position, the mean time and standard deviation for each position within a run, and the number of runs of any given length.

The Low Bias category for the 68 per cent. was identical to the High Bias category for the 44 per cent. group and contained seven trials (54–58 per cent.). For 68 per cent., the Medium Bias category contained 53 (60–68 per cent.), the High Bias, 39 trials (70–80 per cent.)—(one remaining trial of 50 per cent. was not included in the detailed analysis). For 44 per cent., the Medium Bias category contained 55 trials (42–52 per cent.) and the Low Bias category 38 trials (22–40 per cent.), making a total of 100 trials for each programme. The EDSAC analysis identified the location and nature of errors made by each subject. In view of the low overall error percentage which averaged just around one per cent., the major analysis treated errors as if they had been correct responses. Although this caused some distortion of the data, it is small compared with the very large number of entries, 5,000 responses for each subject.

Run analysis

There are two ways of doing this: one can either pick out all runs of equal length and obtain means for each position per subject per run-length. In this case there will be an equal number of entries for each datum point. Alternatively, one can superimpose all runs and obtain mean times for each position per subject. In this case there will be an unequal number of entries for each datum point, more for the earlier positions than for the later ones. Preliminary analysis had shown that adopting the second method did not produce misleading results and since it is by far the easier of the two to programme, it was adopted for most of the data to be presented here.

Results

Input characteristics

The overall structure of the input programmes for the second week is shown in Tables I and II. Table I shows the relative frequencies and percentages with which the five stimuli were found in the two programmes for the second week. It will be seen that as far as signal three is concerned the observed frequencies are very close to the overall bias expected as the result of the generating procedures.

In Table II we show incidence of runs of stimulus three found in the two programmes for the second week. From this it will be seen that there was a higher incidence of short runs for the 44 per cent. programme but that there was a cross over after a run-length of 3 so that there were then more long run-length for the 68 per cent. programme.

TABLE I

RELATIVE FREQUENCIES AND PERCENTAGES FOR STIMULI OBSERVED IN THE INPUT PROGRAMMES FOR THE SECOND WEEK OF THE EXPERIMENT

	Stimulus 68 per cent.		*Stimulus 44 per cent.*	
		per cent.		*per cent.*
1	454	9·08	702	14·04
2	418	8·36	699	13·98
3	3,358	67·16	2,146	42·92
4	368	7·36	692	13·84
5	402	8·04	761	15·22
Total	5,000		5,000	

TABLE II

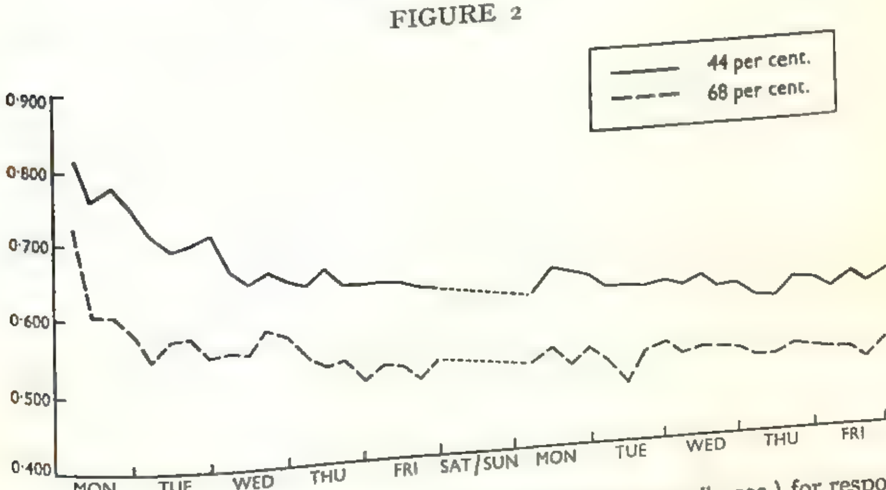
INCIDENCE OF RUNS OF STIMULUS THREE OBSERVED IN THE TWO INPUT PROGRAMMES FOR THE SECOND WEEK

	1	2	3	4	5	6	7	8	9	10
68 per cent.	429	223	174	115	65	50	28	23	14	15
44 per cent.	694	312	131	62	21	8	2	0	0	2
	11	12	13	14	15	16	17	18	19	
68 per cent.	6	5	2	0	2	0	0	0	2	
44 per cent.	0	0	0	0	0	0	0	0	0	

Overall performance comparisons

Figure 2 shows the performance curves for the two groups. The second week's time-scores were summed for each subject and the distributions so obtained compared. There was no overlap so that we can say that the overall difference between the two

FIGURE 2



Performance curves for the two groups in terms of mean time (in sec.) for response for successive blocks.

groups was reliable and the second of the criteria mentioned in the introduction was met.

There were just below 1 per cent. of errors on average for the 68 per cent. group, and just above 1 per cent. for the 44 per cent. group.

TABLE III

MEAN TIMES (IN SEC.) PER RESPONSE TO EACH OF THE FIVE STIMULI FOR THE TWO GROUPS AND THE THREE CATEGORIES WITHIN EACH GROUP FOR THE SECOND WEEK

Average Bias on Stimulus three = 67 per cent.

	1	2	3	4	5
High (73 per cent.) ..	0·636	0·664	0·431	0·633	0·661
Medium (65 per cent.) ..	0·620	0·671	0·445	0·655	0·658
Low (56 per cent.) ..	0·636	0·664	0·462	0·666	0·658

Average Bias on Stimulus three = 43 per cent.

	1	2	3	4	5
High (56 per cent.) ..	0·597	0·627	0·523	0·674	0·650
Medium (46 per cent.) ..	0·605	0·658	0·548	0·670	0·646
Low (36 per cent.) ..	0·612	0·639	0·564	0·668	0·650

Comparisons between categories within groups

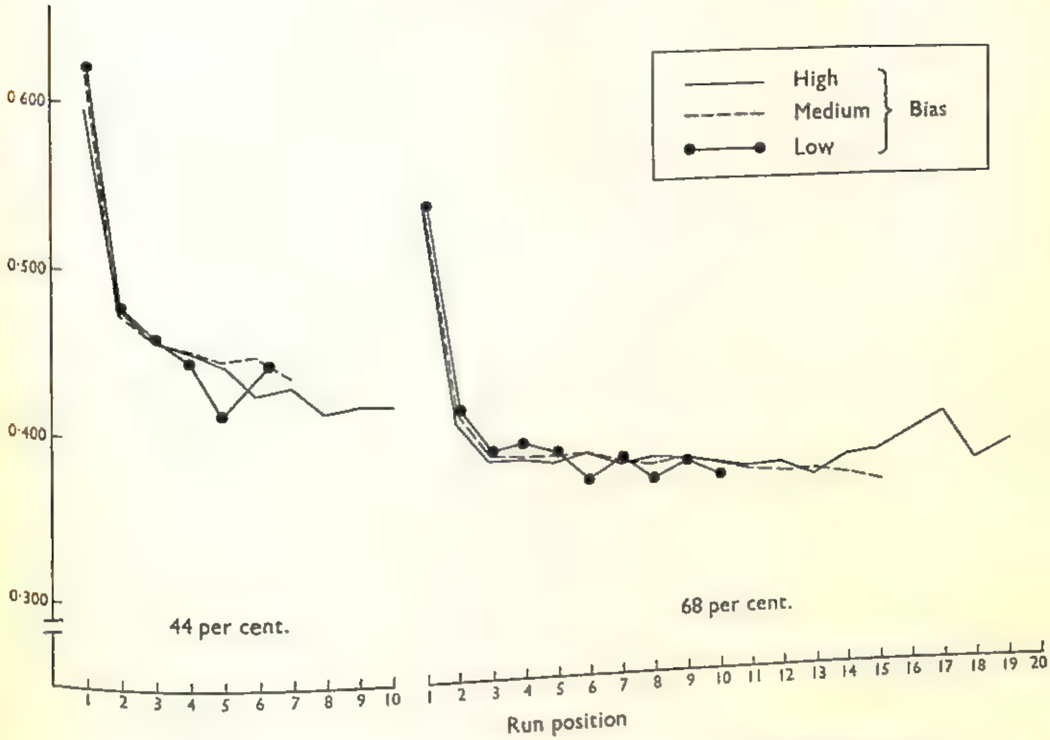
Table III shows the mean times per response to each of the five stimuli for the two groups and the three local bias categories (second week's data). It will be seen that in all instances the response to stimulus three was faster than any of the other stimuli, and this was found to be the case for the means of all individuals. This effect was therefore also reliable. Both groups treated stimulus three as a high probability event, though there was a difference between the two groups (see below). The third of our criteria was thus also met.

It should be noted that within each group there was a decrease in response time to stimulus three from the Low to the High categories. This decrease was found to be present for all subjects in both groups.

Figure 3 shows the results for the second week of the run analyses for the two groups and for the three categories: on the left are the three curves for 44 per cent. and on the right for 68 per cent. In both groups there was a strong "repetition" effect in the sense that the response time to stimulus three appearing in the first position was markedly longer than in the subsequent positions (an effect found to hold for all subjects), and that there was a strong tendency for response times to level out after about the third position. It will further be noticed that while there was little difference in the curves within each group, those for the 44 per cent. group were all on a higher level.

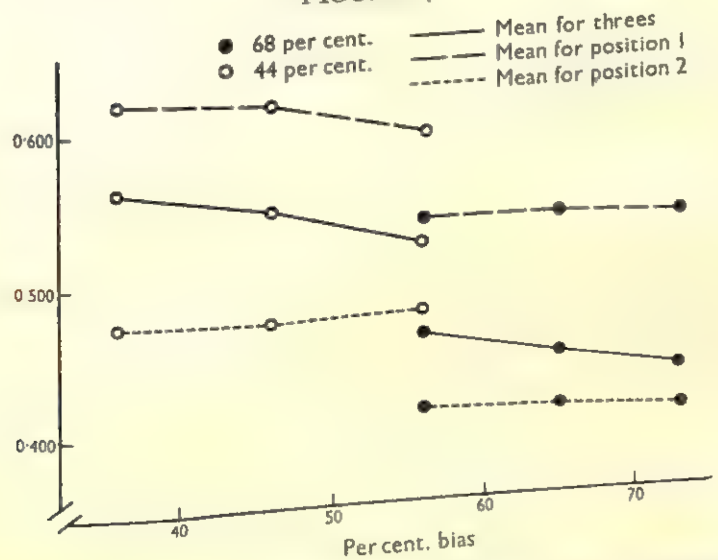
Figure 4 shows the mean times for responses to all stimulus threes (as in Table III above) compared here with their response times to stimulus three in positions 1 and positions 2. For threes in the first position, the only difference which was reliable within groups was that between High and Medium and High and Low for the 44 per cent. For threes in the second position only the difference between Medium

FIGURE 3



Run analyses for the two groups. Mean times (in sec.) for each position.

FIGURE 4



Mean times (in sec.) to all stimuli three, stimuli three in position 1, and stimuli three in position 2 for the two groups and in the three categories of local bias.

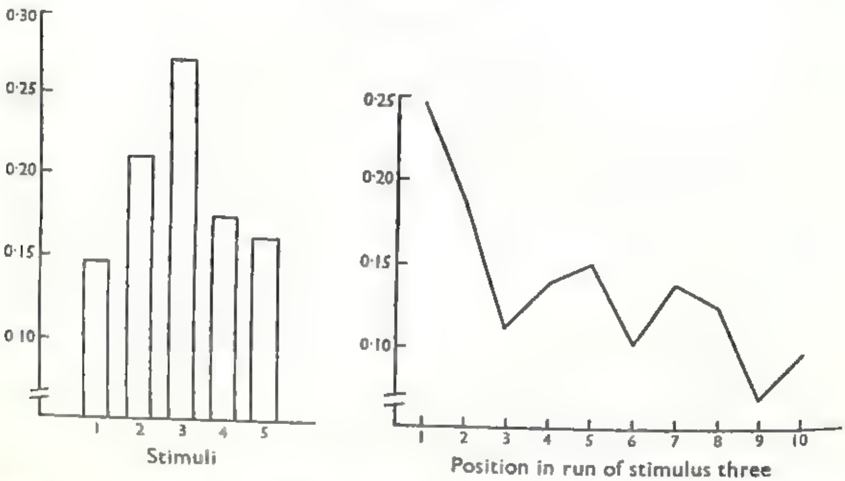
and Low categories for 68 per cent. was shown by all subjects in that group. These differences were, however, very much smaller than those appearing in the mean curve for all threes; the effect of trial bias variation is thus very much less than the "all three" curves suggest.

To sum up the findings so far: we were able to show a strong repetition effect and we were able to show that the response times to "all threes" varied with amount of local bias. But this variation is attributable to the differing proportions of threes in the initial, and later, positions in a run, and not to a change in the response time to threes in equivalent run positions.

Variability

It will be recalled that we obtained standard deviations from the EDSAC print out for all the means obtained. Examination of these standard deviations revealed a very consistent pattern. The standard deviation of the means for responses to stimulus three in position 1 were comparable to those for the other four stimuli, i.e. the non-threes; but the standard deviations of means for successive positions within runs of threes reduced considerably. Beyond the third or fourth positions the standard deviations reduced by a greater amount than the corresponding means. To illustrate this point we are showing the coefficients of variation (s.d./m.) averaged for the Medium category of the 68 per cent. group up to run-length 10, and for the five stimuli separately in Figure 5. All five subjects showed this effect, and it was just as consistent, and marked, in the remainder of the data.

FIGURE 5



Coefficients of variation. Example taken from the medium category of the 68 per cent. group showing the *average* ratios of standard deviations to means for the five stimuli and for the first ten positions of the run-analysis for this sample. Note that the coefficient for No. 3 is particularly high: this is because it includes threes in all positions of a run as shown in the second part of the figure.

Run analysis conclusions

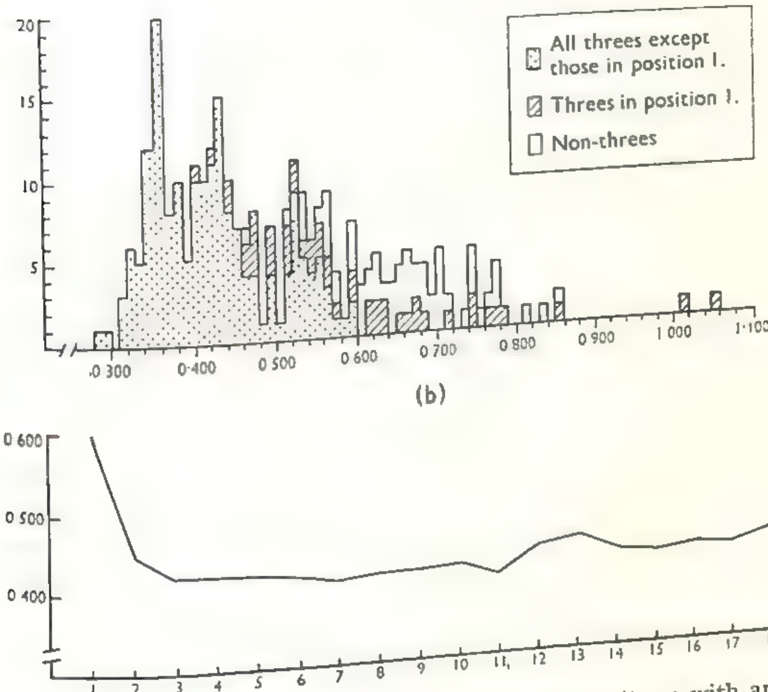
From the data presented so far it would appear that the response times to position 1 were similar to "ordinary" choice reaction times, but that response times to successive positions appeared to be different both in terms of means and of variability. As a final illustration for this point, we have prepared Figure 6 from

the analysis obtained from a pilot group run under 68 per cent. The upper half of this figure is a histogram of all the 300 response times (6 subjects × 50 responses for this one sample trial having 40 threes). The solid portion are responses to threes other than first responses, while the shaded are first responses and the blank portion response times to non-threes. From this histogram, it will be seen that the long end is composed largely of responses to either 1 positions or non-threes. The lower half

FIGURE 6

(a)

(b)



(*a*) Distribution of all responses from a single trial of 50 items with an incidence of 40 stimuli three (of which eight were in the first position) and ten non-threes. Data pooled for six subjects from the group used for the preliminary experiment.

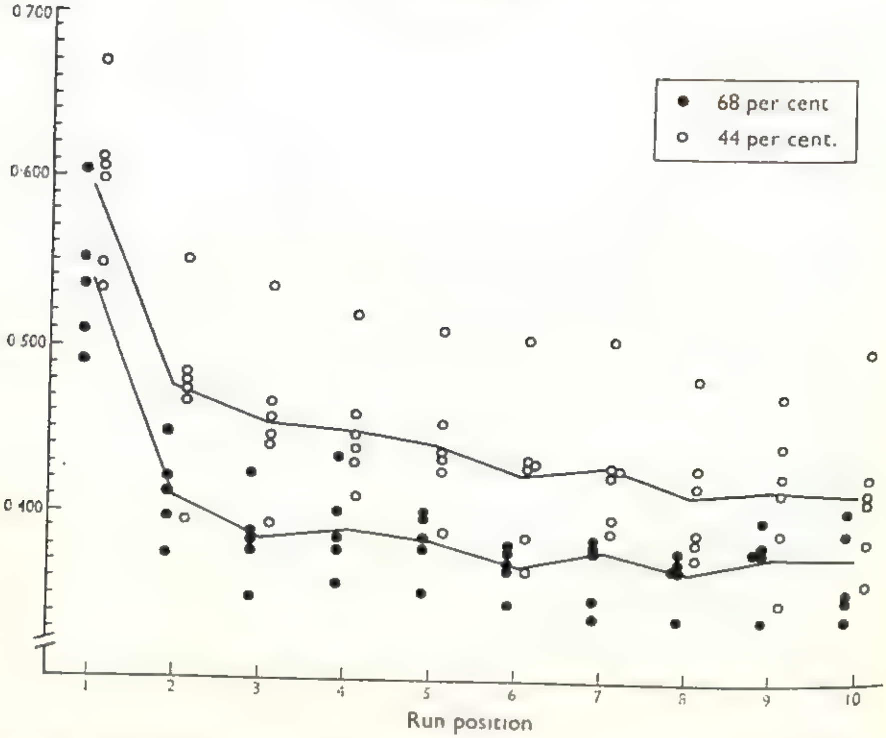
(*b*) Run analysis for the entries shown in (*a*) above. Mean times (in sec.) for successive positions in the run.

of this figure shows the run analysis in terms of mean times for this particular trial. It appears from this comparison that one would be quite unjustified in disregarding the long end of the distribution of all threes since the short end could not be generated without the existence of the readings contributing to the long end!

Analysis of performance of the two groups in the seven identical trials (see Stimulus Sequences above)

The overall mean times per subject for the seven identical trials were not reliably different. There was, however, a reliable difference in the mean response times to all stimuli three and for most of the differences for the 10 run positions except for the first.

FIGURE 7



Run analysis for the seven trials identical in the two groups showing mean times (in sec.) per position for individual subjects.

In Figure 7, we show again two mean curves presented in another context (Fig. 3 above), the mean times per positions in runs for the identical trials, shown separately for the two groups and this time with individual subjects' means plotted as well. This figure shows the similarities and differences between the performance for the two groups in some detail: while there was still a consistent tendency for the response times for successive positions to become reduced in the case of the 44 per cent. group, the curve for the 68 per cent. group was effectively flat from about the third position onwards.

It is clear from this figure and the other data presented that even though overall performance for these identical trials was the same, the constituent times and the trends in the run analysis differed. It is perhaps of some methodological interest to note that one has here an example of the limits of usefulness of overall measures which in this case might have led one to a conclusion rather different from that to be drawn by us. It is also worth noting that the mean times for all the non-threes summed were not reliably slower for the 68 per cent. group than for the 44 per cent. group.

Discussion

Our main findings would seem to be quite unambiguous: to the extent that we have confirmed the existence of Bertelson's repetition effect, we have provided an illustration of "short-term sampling" by subjects under both conditions of the experiment, and at the three levels of local bias examined within each condition. On the other hand, there would seem to be overwhelming evidence for the role of

"long-term sampling" as illustrated by the fact that the pattern of responding to runs was identical within each condition regardless of the level of bias examined and more specifically, by the results obtained from the comparison of the "identical trials." Under the conditions of our experiment, subjects clearly adopted a relatively stable pattern of performance which was unaffected by moment to moment fluctuations. Our results suggest that subjects treated the first appearance of a biased stimulus very much in the same way as the first appearance of any other stimulus, that there was a rapid decrease in response time up to about the third repetition, and that there was no further decrease beyond that. The first entry on what we may call the "repetition" function would therefore seem to be fixed by the response time required for the low-probability stimuli, its subsequent shape by expectancies based on the average probability structure of the input. Subjects adopt a mode of response for dealing with runs which we propose to call "controlled tapping."

The observation that there was a negative correlation between times per block of trial and amount of local bias (Leonard and Newman, 1965) can be accounted for in terms of a slightly more complex additive model than originally expected. The complication is due to the nature of the "repetition function": trials with low local bias will not only have fewer responses to biased stimuli to contribute to the total time, but these few will be drawn from the early and slower part of the function rather than from the later and faster part.

On variability

In Figure 5 we have shown an example of the manner in which relative variability decreased as run length increased, and the extent to which the variability associated with response to position 1 was similar to that associated with the responses to the non-biased stimuli. It is largely because of this reduced variability that we are tempted to propose the term "controlled tapping."

There are many problems which have to be solved before one can tackle a general model. Of these problems perhaps the most interesting will be a study of the early learning process. Most of our own evidence and that of others leads one to suspect that a form of repetition effect can be found quite early on in learning and it would be interesting to see what factors contribute towards its establishment and the manner in which it develops to the form found by us.

REFERENCES

Bertelson, P. (1961). Sequential redundancy and speed in a serial two-choice responding task. *Quart. J. exp. Psychol.*, **13**, 90–102.

Carpenter, A., and Leonard, J. A. (1960). The use of small filaments as visual stimuli. *Quart. J. exp. Psychol.*, **12**, 249–51.

Crossman, E. R. F. W. (1953). Entropy and choice time: the effect of frequency unbalance on choice responses. *Quart J. exp. Psychol.*, **5**, 41–51.

Falmagne, J. C. (1965). A linear model for choice R.T. with application to experimental results. *J. math. Psychol.*, **2**, 77–127.

Hyman, R. (1953). Stimulus information as a determinant of reaction time. *J. exp. Psychol.*, **45**, 188–97.

Leonard, J. A. (1959). Five choice serial reaction apparatus. *Med. Res. Council Rep.*, *APU* 326.

Leonard, J. A., and Newman, R. C. (1965). On the acquisition and maintenance of high speed and high accuracy in a keyboard task. *Ergonomics*, **8**, 281–304.

Welford, N. T. (1952). An electronic digital recording machine, the Setar. *J. Scient. Instr.*, **29**, 1–4.

Manuscript received 23rd November, 1965.

COMPARISON OF THE EFFECT OF HARD AND EASY MENTAL ARITHMETIC UPON BLOCKING OF THE OCCIPITAL ALPHA RHYTHM

BY

ALAN GLASS

From the Department of Anatomy, The Medical School, University of Birmingham

The intensity of blocking of the alpha rhythm of the EEG induced by the solution of five Hard mental multiplications has been compared with the intensity of blocking induced by five Easy mental multiplications in 36 normal subjects. From predictions derived from the application of two hypotheses concerning the intensity of blocking, it would be anticipated, firstly, that blocking would be more intense in the Hard tasks, if their solution required increased mental concentration and, secondly, that insofar as "higher thought" processes are more extensively required in the solution of Hard tasks, then blocking would be less intense in Hard tasks than in Easy tasks. However, the findings of the experiment have been that blocking is of the same intensity in tasks of both kinds. Possible reasons for this are discussed.

Hard tasks differed from Easy tasks, with regard to the relation of EEG changes and performance. It is thought that different mechanisms may underly the solution of the two types of task.

Introduction

Blocking of the occipital alpha rhythm may be a manifestation of either general alertness (Darrow, 1946; Lindsley, 1952) or specifically visual alertness (Adrian, 1944; Oswald, 1962). Blocking induced by mental arithmetic, provides a means of more closely relating the EEG to mental function. The hypothesis that efficient arithmetical performance is controlled by fluctuations in EEG activity has been supported by analysis of the relation of blocking to the answers made to a set of mental multiplications by 58 normal subjects (Glass, 1959, 1964*a*) using a method of measuring the intensity of blocking based on the rate of change of potential (r.c.p.). It was found (*a*) that a high probability of error was significantly correlated with a high r.c.p. before the task or during its solution; (*b*) that a subject's average latency of response was correlated with the intensity of blocking when age was corrected for; and (*c*) that intense blocking was associated with an elevated motivational state. In the present study, this analysis has been extended by examination of the effects of variation in the magnitude of such tasks upon the intensity of blocking which was assessed by the same method. Because blocking is intensified by mental concentration (Adrian and Matthews, 1934; Ostow, 1950), it would be expected that more difficult tasks would induce intense blocking, assuming that greater attention is required for their correct solution. Conversely, blocking may be less intense when associated with tasks of greater difficulty, if, as Costello and McGregor (1957) have suggested, blocking is diminished when "higher thought" processes predominate in the calculations which evoke it. An attempt has been made to resolve these conflicting hypotheses by finding whether both intensity of blocking and its relation to performance were influenced by variation in the magnitude of the calculations.

Correctness and latency of answers and the intensity of blocking were compared in 36 normal subjects. While their EEG's were recorded, they were presented with mental multiplications arranged in two groups of contrasting difficulty. A preliminary account of the results has been given (Glass, 1964*b*).

Method

The subjects reclined, with eyes closed, in a sound-insulated room and were presented with a tape-recorded set of five Hard and five Easy mental multiplications at intervals of 2 min. EEG's were taken as the required calculations were performed and an independent observer noted the answers. An initial EEG recording permitted the subjects to become accustomed to their environment. The subjects were Birmingham medical students and members of staff described in Series A2 (Glass, 1964*a*).

Electroencephalographic recording

Four silver, silver-chloride scalp electrodes were attached in parieto-central and occipital positions corresponding to positions O 1, O 2, C 3 and C 4 of the International Ten Twenty electrode system (International Federation, 1958). A four channel Ediswan machine was used to make bipolar recordings. The four channels connected were C 3—O 1, O 1—O 2, O 2—C 4 and C 3—C 4. Although these are bipolar recordings, the channels C 3—O 1 and O 2—C 4 are an approximation to monopolar recordings because the electrodes were aligned along the voltage gradient of the alpha rhythm (Surwillo, 1963).

Multiplications

Subjects were asked to multiply a two-digit multiplier and a single digit multiplier. In Hard tasks these were composed of digits above five (for example, 68×7) and in Easy tasks of digits below five (for example, 32×4). Five and zero were excluded from the digits used in all tasks. Hard tasks invariably required the "carrying" of the product of the first digit to the product of the second digit. Randomly selected positions 3, 4, 7, 9 and 10 were occupied by Hard tasks and positions 1, 2, 5, 6 and 8 were occupied by Easy tasks in the tape-recorded set.

Method of measuring intensity of blocking

The method of measuring intensity of blocking, which was carried out chiefly in channels C 3—O 1 and O 2—C 4, resembles Jasper and Cruikshank's (1937) "mean deflection" method. The total length traced by the recording pen for a given duration, was found by a map-measuring instrument or opisometer and divided by the direct length representing this duration. The ratio so obtained has been shown to be a function of the average rate of change of potential (r.c.p.), independent of direction of change (Glass, 1964*a*). The r.c.p. in microvolts per sec. was computed by the appropriate formula from this ratio, the time-scale and calibration pulse, in the trace in the 10 sec. prior to the presentation of the task and also in the trace for the period of time during which the calculation was performed.

Index I of the intensity of blocking was calculated by comparing each blocked r.c.p. with the average pre-task r.c.p. Index II of blocking was computed from the average of 10 blocked r.c.p.'s for each subject divided by the pre-response r.c.p. for each task. Index III was the ratio of each blocked r.c.p. to each pre-response r.c.p.

The indices were selected with the object of minimizing the effect of large variations that occur in r.c.p. between subjects. Index I indicated variations in pre-response r.c.p. for each subject, Index II variations in pre-task r.c.p. for each subject, and Index III the ratio between pre-response and pre-task r.c.p. in the solution of each task by each subject.

Statistical methods

Since the data are stratified according to two attributes, subjects and tasks, a two-way analysis of variance was performed. An example of this is given in Table II for latencies of response. The variance between tasks has been divided into that part due to the difference between Hard and Easy tasks and that part due to differences within these two groups of tasks. The first order interactions between subjects and tasks is possibly significant but as there are no replications of the experiments, there is no way of testing this hypothesis. However, the purpose of the analysis has been to determine whether the difference observed in respect of the variables in the two categories of task would be repeated in other (fresh) subjects and so the relevant error terms for testing significance are the subjects-tasks interaction. Correlation coefficients (see Appendix) were calculated between indices, errors and latencies for subjects, tasks and the interaction between them, in both Hard and Easy tasks (Fisher, 1954).

Results

Differences between Easy and Hard tasks

As can be seen from Table I the mean number of errors made by 36 subjects in the five Hard tasks was 0·411 errors per task, but the mean number of errors made in the five Easy tasks was 0·239 errors per task. The average latency of response

TABLE I

Means and Differences of Errors/Task, Latencies and Indices for Hard (H) and Easy (E) Tasks

Type of task	*Errors/task*	*Latency*	*Index I*	*Index II*	*Index III*
H	0·411	19·477	0·739	0·774	0·764
E	0·239	9·673	0·750	0·762	0·758
Difference: (H–E) ..	0·172	9·804	−0·011	0·012	0·006

TABLE II

Analysis of Variance: Latency*

Source of variation	*S.S.*	*d.f.*	*M.S.*	*F.*	*p.*
S = Subjects	12·411	35	0·355	—	—
B = Between H and E	8·327	1	8·327	178·300	$<$0·001
WH = Within H ..	1·076	4	0·269	9·509	$<$0·001
WE = Within E	0·773	4	0·193	6·137	$<$0·001
Interaction					
S × B	1·634	35	0·047	1·562	$<$0·05
S × WH	3·965	140	0·028	—	—
S × WE	4·405	140	0·032	—	—
Total	32·592	359	—	—	—

* These data were logarithmically transformed.

TABLE III

Analysis of Variance

Source of variation	*d.f.*	*Index I*			*Index II*			*Index III*		
		S.S.	*M.S.*	*F.*	*S.S.*	*M.S.*	*F.*	*S.S.*	*M.S.*	*F.*
S = Subjects	35	10·210	0·292	—	10·640	0·304	—	10·140	0·290	—
B = Between H and E	1	0·012	0·012	0·40†	0·010	0·010	0·50†	0·006	0·006	0·12†
WH = Within H ..	4	0·054	0·014	0·64†	0·090	0·022	1·22†	0·110	0·028	0·80†
WE = Within E ..	4	0·477	0·119	4·02*	0·490	0·123	5·86**	0·080	0·020	0·56†
Interaction										
S × B	35	1·040	0·030	1·17†	0·696	0·020	1·00†	1·739	0·050	1·39††
S × WH	140	2·937	0·021	—	2·570	0·018	—	4·950	0·035	—
S × WE	140	4·139	0·030	—	2·990	0·021	—	5·080	0·036	—
Total	359	18·869	—	—	17·486	—	—	22·105	—	—

† $p > 0·2$; †† $p > 0·1$; * $p < 0·01$; ** $p < 0·001$.

was 19·477 sec. in Hard tasks and 9·673 in Easy tasks. This difference was statistically significant whether significance was tested between Hard and Easy tasks ($p < 0{\cdot}001$) or in subjects by Hard and Easy Interaction ($p < 0{\cdot}05$) (see Table II). In contrast, the means of the three indices of blocking were between 0·774 and 0·739 for both Hard and Easy tasks. As might be expected these minor differences were not significant. Analyses of variance for the three indices of blocking are shown in Table III; the only significant variance found was within the Easy tasks for Indices I and II. As Table I shows, the mean of Index III was remarkably consistent in Hard and Easy tasks.

Relation between blocking and performance

Easy tasks. Table IV shows that for Easy tasks, Index I was strongly and significantly ($p < 0{\cdot}01$) correlated with errors, implying that an elevated mean r.c.p.

TABLE IV

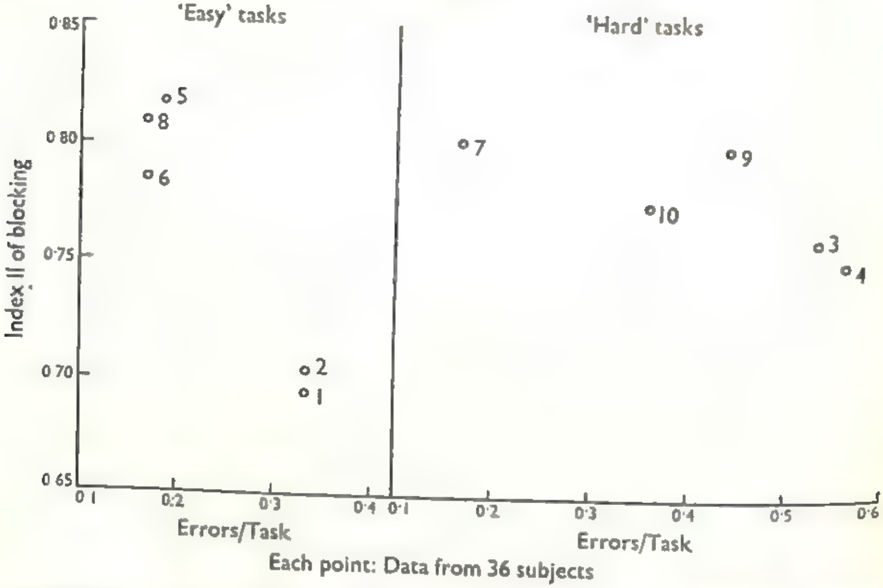
CORRELATION COEFFICIENTS BETWEEN INDICES I, II AND III AND ERRORS AND LATENCIES
EASY TASKS

		Index I		*Index II*		*Index III*	
		r	*p*	*r*	*p*	*r*	*p*
Subjects	Errors	0·199	>0·1	0·167	>0·1	0·214	>0·1
	Latencies	0·195	>0·1	0·193	>0·1	0·165	>0·1
Tasks	Errors	0·958	<0·01	−0·973	<0·01	−0·102	>0·1
	Latencies	−0·640	>0·1	0·897	<0·05	0·770	>0·1
Interaction	Errors	0·031	>0·1	−0·081	>0·1	−0·048	>0·1
	Latencies	0·090	>0·1	−0·107	>0·1	0·002	>0·1
Total	Errors	0·137	>0·1	−0·005	>0·1	0·065	>0·1
	Latencies	0·110	>0·1	0·080	>0·1	0·087	>0·1

Coefficients of correlation (r) are tabulated between Indices I, II and III and errors and latencies of response for Easy tasks. The significance of these differences is indicated by their probabilities (p). The table shows a statistically significant correlation between indices and errors in respect of tasks.

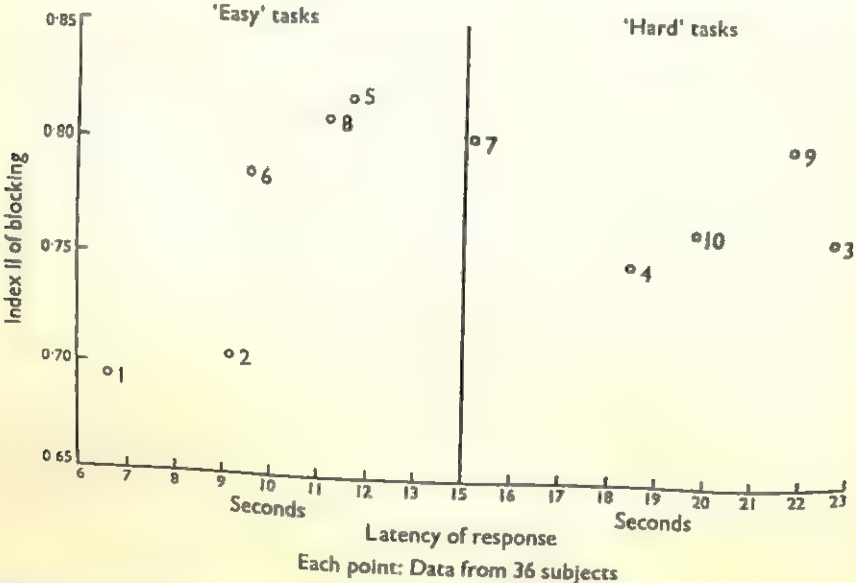
during blocking to Easy tasks increased the number of associated incorrect responses. Index II also was negatively correlated with errors (Fig. 1) at a statistically significant level ($p < 0{\cdot}01$). This is an indication that an elevated mean r.c.p. before the presentation of Easy tasks was associated with a concomitant increase in the number of errors in the answers given. As this was a significant correlation over tasks but not over subjects, it implies that this relation between errors and latencies and Indices of blocking was valid for the particular subjects of this experiment but it may not be applicable to any other group of subjects. As is demonstrated in Figure 2 there was a correlation between Index II and the latency of response for

FIGURE 1



The two graphs compare the relation between Index II and errors per task in five Easy and five Hard tasks. Each point represents data for 36 subjects. Numbers refer to the position of the tasks in the sequence of the experiment. There is a negative correlation ($r = -0{\cdot}973$) for Easy tasks between Index II and errors which is significant whereas in Hard tasks the correlation is smaller and non-significant ($r = -0{\cdot}693$).

FIGURE 2



These two graphs compare the relation of Index II of blocking to latency of response in five Hard and five Easy tasks in 36 subjects. They demonstrate a strong correlation ($r = 0{\cdot}897$) for Easy tasks, which was just significant, but a weak correlation in Hard tasks.

Easy tasks which was just significant ($p < 0.05$). This indicates that the lower the average r.c.p. prior to an Easy task, the longer was the latency of response to that particular task. As this was a correlation residing between tasks, it was probably dependent upon the order of the Easy tasks in the recorded sequence, rather than upon a direct association between a low pre-task r.c.p. and a subsequent long latency of response, because there were significant correlation coefficients which have been calculated but are not tabulated, between order of tasks and Index II and average latencies of response for tasks. Other coefficients of correlation between order of tasks and the remaining factors were calculated and found to be not significant.

Hard tasks. In Hard tasks, Table V shows that Index II was negatively correlated with errors in the interaction between tasks and subjects, at a just significant level ($p < 0.05$). From this it may be inferred that a given subject answered incorrectly

TABLE V

CORRELATION COEFFICIENTS BETWEEN INDICES I, II AND III AND ERRORS AND LATENCIES
HARD TASKS

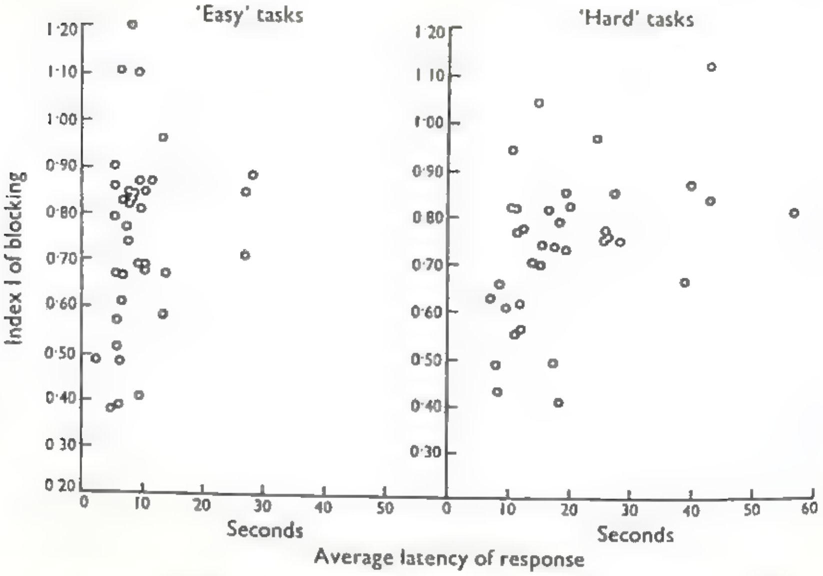
		Index I		*Index II*		*Index III*	
		r	p	r	p	r	p
Subjects	Errors	0·056	>0·1	0·002	>0·1	−0·049	>0·1
	Latencies	0·392	<0·02	0·353	<0·05	0·386	<0·02
Tasks	Errors	0·598	>0·1	−0·693	>0·1	−0·112	>0·1
	Latencies	0·085	>0·1	0·357	>0·1	−0·155	>0·1
Interaction	Errors	−0·037	>0·1	−0·160	<0·05	−0·161	<0·05
	Latencies	0·210	<0·05	0·068	>0·1	0·170	<0·05
Total	Errors	0·023	>0·1	−0·084	>0·1	−0·108	>0·1
	Latencies	0·314	<0·01	0·239	<0·02	0·280	<0·01

As in the previous table but correlation coefficients are tabulated for Hard tasks. Significant correlations are shown between latencies of response and Indices for subjects and a negative correlation of errors with Index II and a positive correlation of Index I with latency occurs in the interaction between subjects and tasks.

more often when high rather than low r.c.p.'s preceded Hard tasks. It should, however, be pointed out, that the precise significance of both the interaction and the total correlations (especially for errors) is not entirely clear and therefore the validity of this conclusion is questionable.

A similar negative correlation of error with Index III ($p < 0.05$) also indicated that for a particular subject, relatively more intense blocking (a decrease in the ratio of pre-response to pre-task r.c.p.) increased the probability of error in solving Hard tasks. This is most likely due to elevated pre-task activity, since the pre-response activity showed no such significant correlation. In Hard (unlike Easy) tasks the average latency of response for subjects exhibited small but statistically significant positive correlations with Indices I, II and III (Table V). Figure 3 shows this

FIGURE 3



These two graphs compare the relation between Index I and latency of response for each subject in Hard and Easy tasks. In Hard tasks the graph shows that there is a positive correlation between Index I and the average latency of response ($r = 0{\cdot}392$) which is significant, but the graph for Easy tasks does not show such a positive correlation.

Each point represents the average of five tasks for each subject.

graphically for Index I. It suggests that a subject's intensity of blocking is related inversely to the calculation time for Hard tasks; the more intense the blocking, the shorter the average calculation-time.

The interaction coefficient of correlation between Index I and latency of response (Table V) was relatively high although only just significant ($p < 0{\cdot}05$). This is an indication that the correlation between the index of blocking and latency of response originates from a high r.c.p. during blocking, since Index I is a function of the variation in r.c.p. during blocking (see Methods). The interaction coefficient for Index III and latency was also just significant, implying that for a given subject, intense blocking is associated with short latencies of response. As shown in Table IV total correlations for the three indices and latencies of response for Hard tasks were statistically significant, due to the significant correlation between latencies and indices for subjects.

Discussion

Direct differences between Hard and Easy tasks

The experiments have shown that although Hard multiplications, containing digits above five, took longer to solve and induced more errors than Easy multiplications containing digits below five, there were no significant differences in intensity of blocking.

When the hypothesis that blocking is intensified when dissociated from "higher thought" processes (Costello and McGregor, 1957), is applied to these results it is assumed that "higher thought" processes would tend to be more extensively employed in the solution of arithmetical tasks of greater than of lesser magnitude (that is, it is assumed that the type of mental processes used in the solution of the

Hard as opposed to the Easy tasks involve some qualitative differences but do not differ markedly in kind). It would therefore be expected that Easy tasks, insofar as "higher thought" processes are less extensively employed in their solution, might induce more intense blocking. But, in fact, the same intensities of blocking were found in both categories of task. From this it may be inferred on first examination that either the assumption itself was not correct or that the results do not support the hypothesis.

The assumption that greater mental concentration would be required to solve Hard tasks might be justifiable because difficulty of arithmetical tasks as measured by calculation-time (latency of response) can be correlated with the logarithm of the factors and their products (Thomas, 1963*a*,*b*). Therefore, applying the hypothesis which associates intensified blocking with increased mental concentration (Adrian and Matthews, 1934), it would be predicted that blocking would be intensified when associated with Hard tasks. From the results, a possible conclusion is that variation in intensity of mental concentration within these limits is without measurable effect upon the intensity of blocking. This tentative conclusion may be contrasted with observations that blocking results from concentrated mental effort (Davis and Davis, 1936; Jasper and Cruikshank, 1937; Travis and Knott, 1936) and that less blocking occurs during the performance of easy non-arithmetical tasks than of harder arithmetical tasks (Chapman, Armington and Bragdon, 1962) but it is in agreement with the finding of only very slight quantitative increase in intensity of EEG blocking during the performance of harder word-opposite tasks than during the performance of easier word-opposite tasks (Martinson, 1939).

Another possible, but unlikely interpretation of these results is that they may present evidence tending to support both hypotheses. The explanation for the intensity of blocking being the same in Hard and Easy tasks is that each factor in both hypotheses may be acting in contrary directions upon the intensity of blocking. Thus, the increased "higher thought" processes required for the Hard tasks may reduce the intensities of blocking, which, however, are increased by the intensified mental concentration required for their solution. Thus the net difference between the intensities of blocking in the two sets of calculations may be supposed to be effectively zero, because of the balance of the influence of the two opposing processes on blocking.

Indirect differences

Errors. It may be inferred from the correlation of pre-response and pre-task r.c.p.'s with errors in Easy but not Hard tasks, that these r.c.p.'s may determine the induction of errors in Easy multiplications which may be "semi-automatic." An important factor in this association may be the recorded sequence of Easy tasks representing the development, in terms of EEG changes, of an increased facility in solving them, which the relatively short duration of each experiment (20 min.) may not have permitted in Hard tasks. Errors in Hard tasks may depend more upon the difficulty of the task. It should however be pointed out that due to the fixed order of presentation of tasks in this experiment, learning and habituation changes may be confounded with differences between Hard and Easy tasks, although the two kinds of task were evenly distributed throughout the sequence and the differences, in some cases, were so marked that it would be difficult to ascribe them to the effect of order.

Negative correlation coefficients for Hard tasks between Index II and III and errors were just significant, indicating that in contrast to Easy tasks for a given individual, an incorrect response to a Hard task was preceded by a high r.c.p. In

calculations of greater magnitude, it would therefore be expected that the correlation between errors and Index II would be increased.

This implies that increased pre-task r.c.p. will influence the occurrence of errors in those tasks, for which, in a given subject, this probability is already high. Responses to tasks which are so difficult that error is almost certain, obviously cannot be affected by changes in pre-task r.c.p. Therefore, it would be predicted that the maximum effect of pre-task r.c.p. will be upon the responses to those tasks of such magnitude which, for the individual, allow an equal probability of either a correct or of an incorrect answer.

These conclusions are in agreement with earlier findings that correct stereognostic pattern recognition (Short and Walter, 1954) and accurate diagram reproduction (Walter and Yeager, 1956) were associated with intense EEG blocking, and the finding of a significant correlation of a subject's logical ability with percent time alpha activity (Beckman and Stein, 1961) and with earlier findings of the relation of performance in mental arithmetic to blocking intensity (Glass, 1964*a*). The degree of difficulty of the arithmetical task which induces blocking must be considered in the interpretation of the relation of efficiency of performance to intensity of blocking. The differences in the relation of errors to intensity of blocking in the two kinds of task would imply that different mechanisms may underly their solution.

Latency of response. For subjects, latency of response was negatively correlated with average intensity of blocking to Hard tasks, which was an indication that individuals who blocked intensely to these tasks also responded quickly to them. In contrast, the average latency of response to Easy tasks was unaffected by the subject's intensity of blocking. Intensity of blocking may therefore influence the latency of response which a subject makes to tasks of the same order of difficulty and of the same order of latency. However, when the differences in latency are great, as in the direct comparison of Hard and Easy tasks, the intensity of blocking is unaffected by the latency of response.

The correlation between latency of response and Index II is just significant for Easy tasks. A short average latency of response to a particular Easy task is therefore associated with a high pre-task r.c.p. which may be an indication of the influence of a sequential learning process, in which a shorter latency of response is associated with a high pre-task r.c.p. in Easy tasks presented later in sequence. Thus, although in Hard tasks, the average latency of response of a subject may be altered by his average intensity of blocking to these tasks (see earlier discussion) the latency of the response to the average Hard task is not affected by the same sequential learning process which would appear to affect latency to the average Easy task.

In Hard tasks, the interaction coefficient between Index I and latency of response is just positive, suggesting that a low r.c.p. during blocking, for a given individual, will be associated with a reduced latency of response to a particular Hard task.

The association of latency of response with intensity of blocking found in the previous study (Glass, 1964*a*) was compared with the association (Surwillo, 1961, 1963) between reaction-time and average alpha frequency which supports the neurophysiological "clock" theory of the alpha rhythm (Wiener, 1958). However, the increased latency of response of Hard tasks compared with the short duration of the reaction time makes it unlikely that the same mechanism can influence both associations.

Conclusions

The finding that the same intensity of blocking is induced by Hard or Easy tasks appears to imply that neither increased mental concentration nor "higher

thought" processes (insofar as the solution of Hard tasks requires the operation of these faculties at a greater intensity than the solution of Easy tasks) can have much influence upon the intensity of blocking of the EEG. However, there is a possibility that the findings may also be explained by the opposing influences on the intensity of blocking of these two processes balancing in Hard and Easy tasks, and resulting in no change of intensity.

In the relations of performance to intensity of blocking, Hard tasks differ from Easy tasks. These differences may indicate that different cerebral mechanisms underly their solution. In Easy tasks, both errors and latency are related to pre-response and pre-task r.c.p., a relation which because of a sequential factor may be associated with learning. In Hard tasks, errors are correlated with pre-task r.c.p. This would lead to the prediction that only in calculations of a magnitude that produced for the individual subject an equal probability of correct or incorrect answers, would elevated pre-task r.c.p. be most strongly associated with errors.

Sir Solly Zuckerman, F.R.S., is thanked for his support and encouragement; Dr. D. Darlington and also Mr. H. Ross are thanked for help in statistics; and to the subjects who participated in these experiments I am especially grateful. This work was supported by the Mental Health Research Fund and the Caroline Harrold Research Fund.

References

Adrian, E. D. (1944). Brain rhythms. *Nature*, **153**, 360–2.

Adrian, E. D., and Matthews, B. H. C. (1934). The Berger rhythm: potential changes from the occipital lobes in man. *Brain*, **57**, 355–85.

Beckman, F. H., and Stein, M. I. (1961). A note on the relationship between per cent. alpha time and efficiency in problem solving. *J. Psychol.*, **51**, 169–72.

Chapman, R. M., Armington, J. C., and Bragdon, H. R. (1962). A quantitative survey of kappa and alpha EEG activity. *Electroenceph. clin. Neurophysiol.*, **14**, 858–68.

Costello, C. G., and McGregor, P. (1957). The relationships between some aspects of visual imagery and the alpha rhythm. *J. ment. Sci.*, **103**, 786–95.

Darrow, C. W. (1946). The electroencephalogram and psychophysiological regulation in the brain. *Amer. J. Psychiat.*, **102**, 791–8.

Davis, H., and Davis, P. A. (1936). Action potentials of the brain in normal persons and in normal states of cerebral activity. *Arch. Neurol. Psychiat.*, **36**, 1214–24.

Fisher, R. A. (1954). *Statistical Methods for Research Workers.* Edinburgh: Oliver Boyd.

Glass, A. (1959). Blocking of the occipital alpha rhythm and problem-solving efficiency. *Electroenceph. clin. Neurophysiol.*, **11**, 605.

Glass, A. (1964*a*). Mental arithmetic and blocking of the occipital alpha rhythm. *Electroenceph. clin. Neurophysiol.*, **16**, 595–603.

Glass, A. (1964*b*). Blocking of the occipital alpha rhythm induced by hard and easy mental multiplication. *Electroenceph. clin. Neurophysiol.*, **16**, 622.

International Federation of Societies for Electroencephalography and Clinical Neurophysiology (1958). Report of the Committee on methods of clinical examination in electroencephalography. *Electroenceph. clin. Neurophysiol.*, **10**, 370–5.

Jasper, H. H., and Cruikshank, R. M. (1937). Electroencephalography: II. Visual stimulation and the after-image as affecting the occipital alpha rhythm. *J. gen. Psychol.*, **17**, 29–48.

Lindsley, D. B. (1952). Psychological phenomena and the electroencephalogram. *Electroenceph. clin. Neurophysiol.*, **4**, 443–56.

Martinson, B. M. (1939). A study of brain potentials during mental blocking. *J. exp. Psychol.*, **24**, 143–56.

Ostow, M. (1950). Psychic function and the electroencephalogram. *Arch. Neurol. Psychiat., Chicago*, **64**, 385–400.

Oswald, I. (1962). *Sleeping and Waking: Physiology and Psychology.* Amsterdam: Elsevier.

SHORT, P. L., and WALTER, W. G. (1954). The relationship between physiological variables and stereognosis. *Electroenceph. clin. Neurophysiol.*, **6**, 29–44.

SURWILLO, W. W. (1961). Frequency of the "alpha" rhythm, reaction time and age. *Nature*, **191**, 823–4.

SURWILLO, W. W. (1963). The relation of simple response time to brain-wave frequency and the effects of age. *Electroenceph. clin. Neurophysiol.*, **15**, 105–14.

THOMAS, H. B. G. (1963*a*). Sources of difficulty in learning arithmetic facts. *Nature*, **199**, 99.

THOMAS, H. B. G. (1963*b*). Communication theory and the constellation hypothesis of calculation. *Quart. J. exp. Psychol.*, **15**, 173–91.

TRAVIS, L. E., and KNOTT, J. R. (1936). Brain potential studies of perseveration time to light. *J. Psychol.*, **3**, 97–100.

WALTER, R. D., and YEAGER, C. L. (1956). Visual imagery and electroencephalographic changes. *Electroenceph. clin. Neurophysiol.*, **8**, 193–9.

WIENER, N. (1958). Time and the science of organisation (first paper). *Scientia (Milano)*, **93**, 199–205.

APPENDIX

CALCULATION OF CORRELATION COEFFICIENTS IN HARD AND EASY TASKS

Three analyses of variance tables were constructed for both Hard and Easy tasks, using sums of squares for each variable and the sums of products. Each correlation coefficient was then calculated by dividing each sum of products by the square root of the product of the sums of squares for each row separately, that is for subjects, tasks, interaction and total. A numerical example of the calculation for errors and Index I in Easy tasks is appended.

ANALYSIS (EASY TASKS)

Source	*D.F.*	*Square root* *S.S.* *Index I* × *S.S.* *Errors*	*Cross-products Index I and Errors*	*r.*
Subjects ..	34	6·853 × 10·228 = 8·372	1·67	1·67/8·372 = 0·199
Tasks ..	3	0·4766 × 1·088 = 0·720	0·69	0·69/0·72 = 0·958
Interaction	141	4·1374 × 20·912 = 9·3016	0·29	0·29/9·3016 = 0·031
Total ..	178	11·467 × 32·728 = 19·373	2·65	2·65/19·373 = 0·137

Manuscript received 3rd January, 1966.

SPECIAL ARTICLE

CENTRAL INTERMITTENCY TWENTY YEARS LATER*

BY

PAUL BERTELSON

From the Laboratoire de Psychologie, Université libre de Bruxelles.

In a few days, it will be exactly 20 years since Kenneth Craik, Fellow of St. John's College, Cambridge, and first Director of the Medical Research Council's Applied Psychology Research Unit, was killed in a street accident, at the age of 31. During the 8 years he had spent in Cambridge, he had offered the most striking example of what a man of exceptional ability can achieve when placed in an exceptional environment. Some of you have known Craik. Sir Frederic Bartlett, who knew him in the way that only close co-operation in a common effort can provide, has given us an outstanding picture of his indomitable energy, of his healthy appetite for knowledge, of his mechanical inventiveness paired with the passion for fundamental issues (Bartlett, 1946). The men of my generation can only watch the influence he has had, not only through his few published writings, but also through the direct impetus he gave to Cambridge research, both applied and fundamental, on human skill.

The objective of this lecture will be to examine the developments up to now of one of the most influential ideas launched by Craik, the idea of intermittency in human sensory-motor activity.

This hypothesis was developed in two short papers, intended presumably for limited circulation and discussion rather than for publication, which Craik had written in March and April, 1945, and which were published posthumously under the title "Theory of the human operator in control systems" (Craik, 1947, 1948).

The papers were concerned mainly with tracking performance on which a part of the applied work done by Craik and his associates during the war had been concentrated. A feature which is common to many practical activities such as driving a vehicle, or aiming with a rifle or a gun at a moving object, is that the function of the operator is constantly to minimize the discrepancy between a target and the part of the situation he controls—the pointer—while an external factor causes either the target or the pointer to move, and so tends to increase the discrepancy. A number of laboratory tasks presenting the same characteristics are available. The early versions used by Craik consisted of a smoked drum kymograph, with a line of which the subject saw only a limited portion through a slit in a screen, and a pointer which he had to keep on the line while this was passing behind the slit.

Craik was applying to this situation concepts derived from the theory of servo-mechanisms. His main thesis was that *man behaves basically as an intermittent correction servo*. This means that, even when the changes in the course are continuous, corrections are made at discrete intervals only. It is opposite to a continuous servo, which at each moment adjusts its response as a function of the discrepancy observed a finite time interval (called the time-lag) before.

* A lecture delivered in Cambridge on 4th May, 1965, under the auspices of St. John's College, and submitted for publication at the request of the Editor. Professor Bertelson was the holder of the Kenneth Craik Research Award in 1964–65 (Ed.).

The evidence for this intermittency was the jerky aspect of records of tracking error. The jerks however are evident only in unpractised subjects. With practice, additional factors, such as anticipation of the future course of the target on the basis of its past behaviour, enter the picture and make the recordings look more continuous. This is why the intermittency was considered as *basic* only.

Careful investigations of the role of anticipation in tracking have since been conducted (Poulton, 1957). They confirm Craik's original suggestion, but also, since the picture which emerges is a rather complex one, they show that continuous tracking was probably not a very good situation in which to study intermittency.

Intermittency in correction can be effected by two classes of phenomena. It can be the product of a threshold mechanism where discrepancies smaller than some value evoke no correction. This interpretation could be dismissed on the ground that speeding up the course, or magnifying its amplitude (two manipulations which increase the speed at which a given discrepancy can be built up), left the periodicity of the corrections unaffected, at a constant value of about two corrections per second. This left only the alternative interpretation, that intermittency originated in the central mechanisms responsible for choosing the response. The idea is that the process of choice is done in discrete units. On the basis of the sensory data available at one moment, one unitary "computing process" is started, using Craik's vocabulary, and "new sensory impulses entering the brain while this central computing process is going on would either disturb it or be hindered from disturbing it by some switching system" (1948, p. 143). If we can make an industrial comparison, the central mechanisms operate like a workshop which starts one unit of fabrication and completes it before it starts the next unit rather than like a continuous assembly line where orders enter continuously at one end and assembled units are delivered continuously at the other end. In the comparison, the orders are the sensory data and the assembled units the responses.

The most frequent interval between corrections was of the order of 0·5 sec., although intervals as short as 0·25 sec. were observed also. Craik put much stress on this two-per-sec. periodicity and, as a result, much of the early work on intermittency concentrated on trying to explain this particular value. Craik himself appears in his papers to be hesitating between two different types of explanations.

The first consisted of postulating an intermittency period of half-a-second. Craik here referred to an experiment published in 1931 by Telford, and which had shown that subjects giving simple key-pressing reactions to auditory signals coming at unpredictable intervals ranging from 0·5 to 4 sec., responded more slowly after a 0·5 sec. interval than after one lasting 1 sec. or longer. Telford had concluded that a voluntary response is followed by a refractory phase, analogous to that long known to exist at the level of simpler physiological systems like nerves or muscles. His own experiments, wrote Craik, showed also a refractory period of about 0·5 sec. This acceptance of the term refractory period by Craik was unfortunate, for the analogy is a very loose one: during the physiological refractory phase, the tissue does not respond at all to a new stimulus, or only to a stronger one, while during the so-called psychological refractory period, a stimulus of same intensity can elicit a response, but with a longer latency. The term however has now gained such wide acceptance that one has to use it.

The second line of explanation consisted of trying to specify what happens during the intermittency period which delays a new response. At one point in his paper, Craik explained that between observation of a misalignment and initiation of a corrective movement, a delay—the "reaction time"—of about 0·3 sec. occurs, and that the correction movement itself takes about 0·2 sec., making a total of 0·5 sec.

These values were taken of course from the step-tracking experiments Craik was conducting with M. Vince. No reason was given why computations for a new correction could not start as soon as the movement was initiated. In fact, later work by Poulton (1954), Leonard (1953) and Crossman (1953) has shown that in many serial skills, overlap of responses with decisions regarding later responses is frequently observed. But no doubt, the suggestions made in the paper were tentative in nature, and we can be pretty confident that Craik would have proceeded very quickly beyond that stage, had he been given the opportunity.

In the years following Craik's death, intermittency remained a purely "Cambridge" problem. These years saw two important developments, which we had better describe before starting any comment on the whole issue.

One was the completion and the publication by Vince (1948) of the experiments she had started under the direction of Craik. These involved a series of step-tracking experiments, where the subject had to follow a course which jumped from one level to another at discrete time intervals. This situation provides a much better test of the intermittency hypothesis than continuous tracking, since here the input consists of sharply defined events, permitting a direct measurement of reaction times and of movement times. When two steps occurred at intervals shorter than about 0·5 sec. the reaction time to the second step was markedly increased. Hick (1948) also published some exploratory experiments where the subject responded to successive signals with different hands, and which showed that the delays in second reactions were not due to motor interference.

The other development of these years is the pursuance by Hick (1948) and by Welford (1952) of the theoretical discussion on the origin of intermittency. Some very similar suggestions were made by the two authors but Welford produced such a clear formulation, that all subsequent work on the subject has been directly inspired by his theory.

The basic idea of Welford's model is directly taken over from Craik. In Welford's words, the delay observed in responding to the second of two closely spaced stimuli is due to "the central processes concerned with two separate stimuli not being able to coexist, so that the data from a stimulus which arrives while the central mechanisms are dealing with data from a previous stimulus have to be 'held in store' until the mechanisms have been cleared" (Welford, 1952, p. 3). With the increasing use by psychologists of the language of communication "central mechanisms" were going to be called the "central channel" and Welford's theory known as "the single channel hypothesis". The important step taken by Welford was to take the reaction time as an estimate of the time during which the central mechanisms were occupied by the first stimulus. If the second stimulus arrives before the end of that time, the reaction time to the second stimulus will be the "normal" reaction time to that stimulus plus the interval separating the arrival of that stimulus from the end of the first reaction time.

This simple principle of no-overlap of adjacent reaction times accounted for delays observed when the second stimulus arrived during the first reaction time. But delays had been observed also for longer inter-stimulus intervals. Postulating that the mechanisms remain busy after the end of the reaction time would have given a much too long prediction for the cases where $I < RT_1$. Welford then adopted Hick's suggestion that further delays could be due to the subject paying attention to his response. More specifically, the central mechanisms could be occupied by stimuli, kinesthetic or other, fed back from the response, which would then create a new intermittency period. In graded movements, such stimuli are more likely to occur at the beginning and at the end of the response movement. An important

principle is that they will cause refractoriness, i.e. occupy the central mechanisms, only if these have not been occupied already by a second external stimulus.

The model thus provides a quantitative prediction of the values of RT_2 to expect after a given RT_1 as a function of the inter-stimulus interval. The function assumes a saw-tooth shape. However, since RT_1 is variable, if we plot *average* RT_2 as a function of the interval, we should expect a smoother curve resulting from the superposition of several of these saw-tooth curves displaced regarding each other along the abscissa. For intervals shorter than the mean RT_1, we should not expect a line decreasing at 45° but a gradually flattening curve. It is a pity that so many investigators have presented their results under the form of averages only. Much stronger tests of the model can be obtained when individual pairs of RTs are considered, as was done by Davis (1956), Welford (1959) and Halliday, Kerr and Elithorn (1960).

Having considered Welford's very explicit formulation, it is easy to see what was so important in the intermittency idea. It was to open the way to an analysis of complex activities, whether apparently continuous or discrete, into basic decision units, each consisting of the choice of the adequate reaction for a particular sample of sensory input. The research programme which immediately followed involved two levels: (*a*) the study of the conditions affecting the main parameters, precision and latency (or reaction time) of the basic decision units, and (*b*) the study of the interactions between successive units. We shall concentrate here on the second part of the programme. About the first, one must remember that a very active movement of research on reaction times has developed in the last 15 years, and has led to a considerable increase of our knowledge. The main stimulus for that development came from the hopes raised by the application of the language of information theory, starting with the pioneer work of Hick (1952) published the same year as Welford's theory, but no doubt another was the fact that students of reaction times had stronger reasons than in the early days of experimental psychology to believe that they were working on something of very general significance.

The idea of "units" of decision may look naïvely atomistic to some, who will see in it nothing but a rephrasing of the old stimulus-response approach. But nothing has been said so far about the size of the units. Craik had some very important comments to make about this. After having stated that the periodicity of corrections was about two per second, he added: "Certainly, it is possible to make successive responses at a greater pace, as in transmitting morse, typewriting, playing the piano, etc., but in such cases it seems as if either the groups of stimuli composing letters of morse, for instance, must have been learned previously and that these groups then become single "stimuli" with which the computing system deals as wholes, or, as in the case of reading music, that the page must be open in front of the player so that he can take in a few notes in succession as one group and respond to them as a unit" (Craik, 1948, p. 147). This of course is closely related to the ideas of Bryan and Harter (1899) on "habit hierarchies" and of Lashley (1951) on movement patterns.

The decision units can thus involve at the input side large samples of sensory data, extending in space or time, and, at the output side, complex patterns of movement. Once again, the intermittency is basic only, but a careful analysis of performance carried out at the correct level must be able to reveal it.

"Grouping" can occur in experiments with successive discrete stimuli also. When two stimuli occur at short intervals, they can be treated as one. In Craik and Vince's step-tracking situation, when an upward jump of the course was followed at 50 or 100 millisec. interval by a return jump of the course, all movement was sometimes cancelled. Welford integrated the grouping principle in his model as a

limiting case, and indicated some conditions under which it can occur. One is distraction: if some stimulus external to the task has been left to occupy the channel, the first stimulus may still be waiting in the peripheral store when the second arrives. The other case is deliberate waiting. In experiments designed to study the refractory period, this is generally avoided by urging the subject to deal immediately with the first stimulus when it comes. But it is well to remember that it is the normal strategy in most serial activities.

The grouping principle is in fact much more than a limiting case. It is an integral part of the intermittency hypothesis. In fact, no really predictive model of intermittency can be built as long as the size of the units to which it applies and the conditions under which they are grouped have not been stated. Unfortunately, in the experimental work on intermittency, grouping has been generally treated as a limiting case, used to explain "apparent exceptions," and has benefited from very little systematic study (except for some exploratory work by Sanders, 1964). This is in my opinion the most serious missing link in the study of intermittency. .

Our task in the remainder of this lecture will be to consider some of the fresh knowledge which has been gained through recent work. This, as we have said already, has been mainly inspired by Welford's theory. It has been conducted mainly with tasks requiring the subject to give successive discrete responses, generally of the traditional key-pressing type, to discrete stimuli. It has concentrated on a few topics which we shall examine in turn.

(1) The reality of delays in responding to a stimulus which arrives during the reaction time to a previous stimulus has been very generally confirmed. Most of the published work has been conducted with pairs of simple reactions, where the subject knows which signals are going to come and which responses he must give; his only uncertainty concerns the time of arrival of the signals (Fraisse, 1957; Davis, 1956, 1957, 1959; Kay and Weiss, 1961). In some experiments (Fraisse, 1957; Kay and Weiss, 1961) both responses were given with the same key. In that situation, the delays can be influenced by the time necessary to release the key before pressing it again. This type of motor interference however makes only a very small contribution, as was shown by Fraisse. The most direct demonstration that the delays originate mainly in a *central* bottleneck has been provided by Davis (1957) in an experiment where on each trial the subject responded first with one key to an auditory signal, and then with another key to a visual one.

To some extent, it is a pity that so much of the work has dealt with simple reactions, since choice reactions are much more representative of decisions taken in most skilled activities. Some authors (Marill, 1957; Elithorn and Lawrence, 1955; Halliday, Kerr and Elithorn, 1960; Elithorn, 1961; Kerr, Mingay and Elithorn, 1965) have used a situation where one of two signals, calling for one of two responses, comes first and is followed by the arrival of the *other* signal, calling for the other response. This situation is very complex, since it involves, beside intermittency, the passage from a choice to a simple reaction, and in the ignorance of the control value against which to evaluate the reaction time to the second signal, the interpretation of the results is very difficult.

Welford (1959) has studied choice reactions to two lamps flashing at irregular intervals, and found delays consistent with his model. In this situation, the same reaction can have to be repeated, and this involves motor interference beside central refractoriness. Delays are in fact somewhat longer in that case. Repetition of the response can be avoided by the use of pairs of independent choice reactions. On each trial, one reaction is first given to one of two signals with one of two keys, then another reaction to one of two other signals with one of two other keys, and the subject

knows in which order the reactions will be called for. Borger (1963) has been the first to study this situation, in an experiment where the same inter-stimulus interval was presented for a run of trials. Rather unreliable delays were obtained. One of the three subjects systematically waited for the second signal, and another one, who does not seem to have been waiting, gave very short delays (of the order of 50 millisec.) at short intervals. One might be tempted to attribute this puzzling result to the use of predictable intervals. But since the publication of Borger's paper, Broadbent and Gregory (personal communication, 1965) and myself (unpublished data) have separately conducted several experiments, involving a total of more than 30 subjects, with a basically similar set-up and also predictable intervals, and found very clear delays. In one of these experiments, I made a direct comparison of predictable and unpredictable intervals, and observed a difference smaller than 25 millisec. in the second reaction times. It is surely unnecessary to assume, as Sanders (1964) has recently done, that the case of choice reactions is different from that of simple reactions. The main difference between Borger's procedure and ours is that he insisted much less than we did on the necessity to process immediately the first signal. His findings would imply that strategies intermediate between successive handling and grouping are available under these conditions, but it is probably safer to wait for some confirmation before speculating on this possibility.

Predictable intervals can only be used with choice reactions. It must be clear of course that, if the subject knows which the second signal will be and and when it will come, he can start organizing a double response on the basis of the first signal so that no delay of central origin must be expected. This was implicit already in Craik's formulation, since in such case computation need not be delayed until the arrival of the second signal, but apparently this point has not been universally understood, for the fact that delays depend on stimulus or time uncertainty has been recently presented as a qualification to the intermittency principle (Adams, 1962; Creamer, 1963).

(2) When the second stimulus comes after the end of the first reaction time, the picture is more confused. Davis (1956) in an experiment with pairs of simple reactions to flashes of light, found no delay in the second reaction when the inter-stimulus interval was longer than the first reaction time. He concluded that occupation of the channel by stimuli fed back from the response is at any rate not unavoidable. Other authors found delays consistent with Welford's feedback hypothesis. The most frequent finding is that some delay still occurs after the beginning of the first response, but too short to fit with Welford's hypothesized value of 150 millisec. as the duration of occupation by feedback stimuli. One possible explanation for these short values would be grouping of feedback stimuli with the second external signal (Welford, 1959). Roy Davis' cautious conclusion that occupation by feedback stimuli is not unavoidable undoubtedly holds. Conditions for occupation are not yet clear. The most likely are accuracy required in the response, and level of practice.

(3) It has probably been realized already that the origin of the refractoriness linked to feed-back stimuli is different from that created by stimuli one has to react to, in that it is occupation of the channel without subsequent reaction. The corresponding phenomenon has been demonstrated in the case of external stimuli. Fraisse (1957) and Davis (1959) have separately shown that a stimulus to which no reaction must be given, can cause a delay in the response to another stimulus coming shortly afterwards. The delays seem to be shorter than in the case where a response must be given, but since the authors give their results only in the form of average RTs, it is possible that this reflects a mixture of cases of full delay and of

cases of no delay. Whatever the explanation, there are also cases where stimuli requiring no response cause no delay (Borger, 1963; Rubinstein, 1964).

What is involved in these experiments is the problem of how the operator deals with irrelevant stimuli. It seems that they can either be allowed to occupy the channel and start an intermittency interval which terminates with the decision not to respond, or be refused access to the channel. The second possibility implies that the "switch" placed by Craik before the computing mechanisms is more than a simple shutter; it has a selective ability, and can discard classes of inputs. This function is of course what psychologists have long been calling "attention".

Attention has been the object of much study (Treisman, 1964) through experiments which are not refractoriness experiments strictly speaking, in the sense that they no not involve reaction time measurements, but which can be considered as extensions of the usual refractoriness experiment. The subject is presented simultaneously or in close succession with several messages, generally verbal, and asked to reproduce them, or give them some sort of discriminative response implying that they have been analysed. Unless the messages are very short, he usually can only respond to part of them. But if he is asked to give priority to one message having some characteristic like being spoken in a female voice or coming from the left, he does much better with that particular one, at the cost of doing worse with the other message. Broadbent (1958) has shown that these selective listening phenomena can be dealt with by an extension of the single channel theory which puts much stress on the properties of the selective switch, which he calls a "filter" and of a peripheral memory where unselected messages can be stored for a very short time. In selective listening experiments, messages which are not selected for immediate passage through the channel are completely lost because their stored trace fades before the channel is cleared. With the simple stimuli used in refractoriness experiments, no substantial delay can take place during the short delays involved, so that second stimuli can usually be retrieved from the store.

There have been other applications of the intermittency idea to psychological operations other than the one it was originally applied to, i.e. the choice of an immediate response. Broadbent (1958) is also responsible for an application to short-term memory, contending that decay of memorized traces can be avoided by circulating them through the single channel. And there are obviously a number of operations for which one can try to evaluate the extent to which they use the channel.

These developments are clearly very important. But although they show the fruitfulness of the intermittency idea, the evidence they provide about it is rather indirect, and always less accurate than that provided by reaction-time experiments. Going back to selective listening, it is often not clear whether the choice of the relevant message is achieved through pure filtering, without any occupation of the channel, or through a short analysis, limited to the identification of the cue labelling the relevant message, but occupying the channel for a short intermittent period. Experiments like Davis's or Rubinstein's give much clearer results from that point of view.

(4) So far, we have considered intermittency at the qualitative level only. If we now turn to the more quantitative aspects, Welford's assumption that the occupation of the channel is equal to the reaction time has been questioned. It is well to remember that this hypothesis has the merit of being very simple and of identifying occupation time with an observable variable, this leading to precise predictions, but that it is not the only one consistent with the intermittency hypothesis. Davis (1957) argued that only a part of the observed reaction time is spent in the central nervous system, and that part is due to peripheral processes. If all that was involved was

no-overlap of central times, the motor peripheral part of the first reaction time and the sensory peripheral part of the second would play no part in determining the delay. From there, Davis was led to suggest that, after the end of the central time necessary to organise the response, a true central recovery time must elapse before a new stimulus can be dealt with. An alternative suggestion was made by Welford (1959): it is that before another stimulus is accepted into the central channel, some kind of signal indicating that the response has begun must be sent back to the central mechanisms; as Welford notes, the hypothesis is for the time being purely speculative but it offers an explanation for the fact that delays are generally shorter after first signals to which no response must be given (Fraisse, 1957; Davis, 1959).

Broadbent (1958), following a suggestion originally put forward by North (1954) has considered a completely different type of intermittence mechanism. In Craik's original formulation, the "switch" was there to protect ongoing computations from interference from new inputs, which implied that it would open again when the computations, plus possibly some clearing processes, were over. In the new version, there would be a minimum time interval between the closing of the switch and its opening again, which would not depend on the length of the operations performed on the accepted input. One attraction of this type of model was that a similar quantization of sensory input has been hypothesized to account for a number of other phenomena (Stroud, 1955; Shallice, 1964). The main difference between this hypothesis and all the previous ones is that it predicts a delay which is independent of the duration of the first reaction time. The prediction has been tested directly by Broadbent and Gregory in an unpublished experiment where they manipulated signal-response relationships, and we have to thank the courtesy of the authors for being allowed to say that the results are opposed to the prediction of their model.

My opinion is that for the time being the most promising sort of theory is one incorporating no-overlap of central times and some additional source of delay. A number of such additional sources offer themselves, and the difficulty is rather to choose between them. We need well-analysed quantitative data collected in a larger range of situations before any decision can be reached.

(5) The discussions considered until now have dealt either with the conditions under which intermittency is observed, or with the exact duration of the intermittency period. The fact of intermittency itself has not been contested.

We have now to consider a thesis, known as the "expectancy theory," which has been formulated several times (Poulton, 1950; Elithorn and Lawrence, 1955; Adams, 1962) and holds that the delays observed in second reaction times are artifacts of the situation in which they were measured. It is based on the well-demonstrated fact that the reaction time to a signal varies with the probability of occurrence of that signal as seen by the subject, or in other words that the higher the expectancy of the subject regarding the occurrence of one particular signal at one particular moment, the shorter the reaction time. The argument has been used to explain the results of experiments, the majority of the published ones, where different inter-stimulus intervals were presented in random order with equal frequencies. Since the second signal must come eventually, the probability that it will come next given that it has not come yet is lowest immediately after presentation of the first stimulus and then increases continuously. The longer reaction times observed with the shorter intervals would only reflect the adaptation of the subject to these changes in immediate probability. In fact, a phenomenon of that sort has been observed several times, although not perfectly reliably, in the situation where a signal to which the subject must respond is preceded at varying intervals by a warning signal. The most frequent finding (Breitwieser, 1911; Karlin, 1959; Drazin, 1961) has been

that the reaction time was longest at short intervals and decreased afterwards. The largest effect observed, however, is of the order of 50 or 60 millisec. which is much less than the delays observed in refractoriness experiments.

This already made the hypothesis rather untenable as an explanation of the delays. More direct tests are also possible. Davis (1965) has measured simple reactions to a signal arriving at variable intervals after the subject pressed a first key. The intervals were the same as in previous experiments by the author with pairs of reactions (Davis, 1957, 1959), so that the expectancy of the subject regarding time of occurrence of the signal must have been the same. Nevertheless, the reaction time was independent of the interval. Another result inconsistent with the expectancy hypothesis is that of an experiment where I compared delays in second reactions after predictable and unpredictable intervals and found practically the same values. On the expectancy hypothesis, predictable intervals would produce no delays.

(6) It may seem strange that the expectancy hypothesis has received so much attention. One reason is that two affirmations of very different strength have been continuously mixed up in the discussion. One, the strong version, is that delays are nothing but expectancy effects. This was always unlikely, and has now been shown to be completely untenable. The other version, the weaker, is that delays are influenced by expectancy regarding the signals to come. Kay and Weiss (1961), for example, have varied separately predictability of the first and the second signal in a dual reaction situation, and shown that any increase in uncertainty increases the observed delays. This is perfectly compatible with a theory of Welford's type. The length of the second reaction time is affected directly by expectancy regarding the second signal, and indirectly by expectancy regarding the first, since the delay is a function of the first reaction time. This weaker version of the expectancy hypothesis thus amounts to reminding us that the intermittency interval is not constant, but varies from moment to moment, as a function of prior knowledge about the signals to come.

(7) Now, this combination of expectancy effects with intermittent decisions poses an interesting problem. How are changes in expectancy achieved, in an intermittent or a continuous way? The question can be asked, for instance, about time expectancy. It is well known that if a signal requiring a response is preceded by a warning signal at a short predictable interval, like 0·25 or 0·5 sec., the reaction time is shorter than with longer or unpredictable intervals (Klemmer, 1956; Karlin, 1959; Bertelson and Boons, 1960; Boons and Bertelson, 1961). This shows that the warning signal is used as a time cue to start some kind of preparatory adjustment. If adjustments were intermittent, the subject could at very short preparatory intervals either pay no attention to the signal so that there would be a reaction time as long as that with long or variable intervals, or accept the signal and have a still longer reaction time. I have done an experiment (unpublished) dealing with that problem. On each trial, the subject first waited a long time (5 sec.), then heard a click which was followed after a foreperiod by the lighting of one of two lamps, and had to respond with one of two keys. The foreperiod was the same for 12 consecutive trials, and took values ranging from 0 to 300 millisec. The results show a continuous decrease of the reaction time with increasing foreperiods up to 100–150 millisec. There is thus no evidence for intermittency.

An exploratory experiment by Leonard (1958) suggests that a similar phenomenon can take place in the case of event expectancy.

These facts lead to a qualification of the intermittency hypothesis in the following way. When a signal calls for immediate reaction, it starts an intermittency period.

But when no immediate action is required, but the signal brings information permitting a better prediction of the nature or time of occurrence of another signal, it can start a continuous adjustive process, during which a signal requiring a response can be accepted at any moment.

Another interpretation, but which also involves a qualification of the intermittency principle, is that decision units can be cut off into smaller parts, when the situation makes that strategy more economical. In the case of our experiment, part of the decisional work would be done during the foreperiod, and then the switch would be opened again to accept the stimulus. The difference between this hypothesis and the former one is that this "adjustable intermittency" can work only when the time at which the second batch of information will arrive is known, so that the second opening of the switch can be programmed. The way to decide between the alternatives is thus to consider what happens with variable short foreperiods.

There are still many points to discuss. From the start, I had to give up the hope to do justice to every important contribution, let alone to be complete. I only hope that some general picture of Craik's great contribution emerges from this complicated and sometimes confused story. May it be the picture not of a *law* to display in the showcases of science, but of a provocative, controversial and, for that reason, fruitful idea. This I think is how Craik would have liked it.

References

ADAMS, J. A. (1962). Test of the hypothesis of psychological refractory period. *J. exp. Psychol.*, **64**, 280–7.

BARTLETT, F. C. (1946). Kenneth J. W. Craik, 1914–1945. *Brit. J. Psychol.*, **37**, 109–15. Reprinted in SHERWOOD, S. (Ed.) (1966). *K. J. W. Craik: The Nature of Psychology.* London: Cambridge University Press.

BREITWIESER, J. V. (1911). Attention and movement in reaction time. *Arch. Psychol.*, **18**, 49 pp.

BERTELSON, P., and BOONS, J. P. (1960). Time uncertainty and choice reaction time. *Nature*, **187**, 531–2.

BOONS, J. P., and BERTELSON, P. (1961). L'influence de l'incertitude temporelle sur le temps de réaction de choix. *Année psychol.*, **61**, 361–76.

BORGER, R. (1963). The refractory period and serial choice reactions. *Quart J. exp. Psychol.*, **15**, 1–12.

BROADBENT, D. E. (1958). *Perception and Communication.* London: Pergamon.

BRYAN, W. L., and HARTER, W. (1899). Studies in the telegraphic language. *Psychol. Rev.*, **6**, 345–75.

CRAIK, K. J. W. (1947). Theory of the human operator in control systems. I. The operator as an engineering system. *Brit. J. Psychol.*, **38**, 56–61.

CRAIK, K. J. W. (1948). Theory of the human operator in control systems. II. Man as an element in a control system. *Brit. J. Psychol.*, **38**, 142–8.

CREAMER, L. R. (1963). Event uncertainty, psychological refractory period, and human data processing. *J. exp. Psychol.*, **66**, 187–94.

CROSSMAN, E. R. F. W. (1953). Entropy and choice time. *Quart. J. exp. Psychol.*, **5**, 41–4.

DAVIS, R. (1956). The limits of the "psychological refractory period." *Quart. J. exp. Psychol.*, **8**, 24–38.

DAVIS, R. (1957). The human operator as a single channel information system. *Quart. J. exp. Psychol.*, **9**, 119–29.

DAVIS, R. (1959). The role of "attention" in the psychological refractory period. *Quart. J. exp. Psychol.*, **11**, 211–20.

DAVIS, R. (1965). Expectancy and intermittency. *Quart. J. exp. Psychol.*, **17**, 75–8.

DRAZIN, D. H. (1961). Effects of foreperiod, foreperiod variability and probability of stimulus occurrence on simple reaction time. *J. exp. Psychol.*, **62**, 43–50.

ELITHORN, A. (1961). Central intermittency: some further observations. *Quart. J. exp. Psychol.*, **13**, 240–7.

Elithorn, A., and Lawrence, C. (1955). Central inhibition: some refractory observations. *Quart. J. exp. Psychol.*, **7**, 116–27.

Fraisse, P. (1957). La période réfractaire psychologique. *Année psychol.*, **57**, 315–28.

Halliday, A. M., Kerr, M., and Elithorn, A. (1960). Grouping of stimuli and apparent exceptions to the psychological refractory period. *Quart. J. exp. Psychol.*, **12**, 72–89.

Hick, W. E. (1948). Discontinuous functioning of the human operator in pursuit tasks. *Quart J. exp. Psychol.*, **1**, 36–51.

Hick, W. E. (1952). On the rate of gain of information. *Quart. J. exp. Psychol.*, **4**, 11–26.

Hick, W. E., and Welford, A. T. (1956). Central inhibition: some refractory observations. Comments on the paper by Alick Elithorn and Catherine Lawrence. *Quart. J. exp. Psychol.*, **8**, 39–41.

Karlin, L. (1959). Reaction time as a function of foreperiod duration and variability. *J. exp. Psychol.*, **58**, 185–91.

Kay, H., and Weiss, A. D. (1961). Relationship between simple and serial reaction time. *Nature*, **191**, 790–1.

Kerr, M., Mingay, R., and Elithorn, A. (1965). Patterns of reaction time responses. *Brit. J. Psychol.*, **56**, 53–9.

Klemmer, E. T. (1956). Time uncertainty in simple reaction time. *J. exp. Psychol.*, **51**, 179–84.

Lashley, K. S. (1951). The problem of serial order in behavior. In Jeffress, L. A. (Ed.), *Cerebral Mechanisms in Behavior*. New York: Wiley, 112–36.

Leonard, J. A. (1953). Advance information in sensory-motor skills. *Quart. J. exp. Psychol.*, **5**, 141–9.

Leonard, J. A. (1958). Partial advance information in a choice reaction task. *Brit. J. Psychol.*, **49**, 89–96.

Marill, T. (1957). The psychological refractory phase. *Brit. J. Psychol.*, **48**, 93–7.

North, J. D. (1954). *The Rational Behaviour of Mechanically Extended Man.* Boulton Paul Aircraft Co.

Poulton, E. C. (1950). Perceptual anticipation and reaction time. *Quart. J. exp. Psychol.*, **2**, 99–113.

Poulton, E. C. (1954). Eye-hand span in simple serial tasks. *J. exp. Psychol.*, **47**, 403–10.

Poulton, E. C. (1957). On prediction in skilled movement. *Psychol. Bull.*, **54**, 467–78.

Rubinstein, L. (1964). Intersensory and intrasensory effects in simple reaction time. *Percept. mot. Skills*, **18**, 159–72.

Sanders, A. F. (1964). Selective strategies in the assimilation of successively presented signals. *Quart. J. exp. Psychol.*, **16**, 368–72.

Shallice, T. (1964). The detection of change and the perceptual moment hypothesis. *Brit. J. stat. Psychol.*, **17**, 113–35.

Stroud, J. M. (1955). The fine structure of psychological time. In Quastler, H. (Ed.), *Information Theory in Psychology*. Glencoe, Ill.: Free Press.

Telford, C. W. (1931). The refractory phase of voluntary and associative responses. *J. exp. Psychol.*, **14**, 1–35.

Treisman, A. M. (1964). Selective attention in man. *Brit. med. Bull.*, **20**, 12–6.

Vince, M. (1948). The intermittency of control movements and the psychological refractory period. *Brit. J. Psychol.*, **38**, 149–57.

Welford, A. T. (1952). The "psychological refractory period" and the timing of high-speed performance—a review and a theory. *Brit. J. Psychol.*, **43**, 2–19.

Welford, A. T. (1959). Evidence of a single channel decision mechanism limiting performance in a serial reaction task. *Quart. J. exp. Psychol.*, **11**, 193–210.

Manuscript received 5th October, 1965.

SHORTER ARTICLES AND NOTES

VISUAL SEARCH WITH MEANINGFUL AND NON-MEANINGFUL MATERIAL

BY

KENNETH T. STRONGMAN and ROBERT BROWN
From the Department of Psychology, University of Exeter

Two experiments are reported which attempt to assess the effects of variations in target word, context items and instructions on performance in a visual search task. In Experiment 1, subjects were required to search through context lists of three-letter nonsense syllables (of either high or low association value) for three-letter meaningful target words (of either high or low frequency). They were given either "positive" or "negative" instructions, i.e. were told either to pick out the meaningful word or to pick out the word which was not a nonsense syllable. The results showed that visual search times were significantly influenced by both frequency of target word and association value of context items. A significant interaction was observed between type of instructions and target word frequency. The design of Experiment 2 followed that of Experiment 1, with the exceptions that nonsense syllables now became target items, and meaningful words formed the contexts. Again, nonsense syllable association value and word frequency were found to be critical in determining visual search times.

Introduction

A visual search is one in which the subject is required to pick out the salient target item from a context of irrelevant items. For example, he may be required to search through a context of letters in order to pick out a number.

Much of the previous work in this area has been concerned with visual search as a highly practised skill, studying a small number of subjects extensively until their performance is optimal (cf. Neisser, 1964*a*). Neisser (1963) has also been concerned with varying the structural properties of the stimulus material (e.g. number of letter in each item, shape of the target letter, etc.). Finally, introductory work has been carried out on the semantic properties of the stimulus material (cf. Neisser, 1964*b*) and it is with a more systematic study of this nature that the present experiments are concerned.

Neisser (1964*b*) observed that it takes longer to search for and find a word which is defined by its class membership (e.g. any animal's name) than it does to look for a specific word. The main purpose of the present study was to investigate the effect of semantic variables at a somewhat more fundamental level. A target was defined as either (*a*) the one meaningful word embedded in a context of nonsense words, or (*b*) the one nonsense word embedded in a context of meaningful words. Meaningful words were taken from the Thorndike-Lorge *Teachers' Wordbook* (1944) and nonsense words from Glaze's (1928) lists. This material had the advantage that word frequency and nonsense syllable association value could be systematically varied.

In a search task with this type of material, the simplest way of conceptualizing the processes involved is to assume (*a*) that each context word must be appraised and discarded, and (*b*) that each target word must be appraised and retained. More specifically, it was thought that the greater the semantic disparity between a target word and its context, the easier would be the search task. Thus, in general, high frequency word (HFW) targets should be detected more quickly than low frequency word (LFW) targets. Further, search performance should be relatively better with low association value nonsense syllable (LANS) contexts than with high association value nonsense syllable (HANS) contexts.

Conversely, a LANS target should be easier to detect than a HANS target, and performance should be relatively better with HFW contexts than with LFW contexts.

To test these predictions, subjects either had to search for a meaningful word in a nonsense context (Experiment 1), or a nonsense word in a context of meaningful items (Experiment 2). Within each experiment, two levels of word frequency and two levels of nonsense syllable association value were varied factorially.

Concerning the structural properties of stimulus material, Neisser (1963) found that it took less time to pick out the one item in a series which *contained* a particular target letter than it did to pick out the one item which *did not contain* a particular letter. In the present study, an attempt was made to ascertain whether or not this type of result would apply to semantic attributes rather than structural ones. Although it is impossible to know whether a subject is searching for an item which is unique in possessing "meaningful" properties, or whether he is merely looking for the item which is "not nonsense," it was thought that these strategies might be influenced by experimental instructions. Extrapolating from Neisser's (1963) findings, it was expected that instructions designed to induce search for an item *not* possessing some attribute (hereafter "negative instructions") would lead to longer response times than those designed to induce search for an item *possessing* some attribute (hereafter "positive instructions"). Therefore, within each experiment, half the subjects were given positive instructions and half negative.

Method

Experimental design

Experiment 1. A 2 × 2 × 2 factorial design was used. The main factors were HFW and LFW targets, HANS and LANS contexts, and positive and negative instructions. Two men and two women were tested under each condition.

The task of each subject was to search through seven lists, each of which contained a target word and nine context syllables. The order of presentation of the lists was randomly varied for each subject.

Experiment 2. The design of Experiment 2 was the same as that of Experiment 1, with the exceptions that the targets were now HANSs or LANSs, and the contexts HFWs or LFWs.

Stimulus materials

Every stimulus item was of the form consonant-vowel-consonant. The HFWs were listed by Thorndike and Lorge as appearing over 100 times per million in written English, and LFWs as appearing less than 10 times per million (proper nouns and non-English words were excluded). HANSs had 100 per cent. association value, and LANSs 0 per cent. association value (Glaze, 1928).

Any one stimulus list was made up of nine context items randomly selected from one of the above pools, and one target word similarly assigned. In the set of lists appropriate to each experimental condition, positions 1–3 remained unused for target words, and positions 4–10 were used once each. All lists were typed in black capitals (Elite). They appeared in a double-spaced vertical column in the centre of white cards measuring 6 in. by 4 in.

Subjects

In each experiment, 16 male and 16 female undergraduates (age range 18 to 36, median 20) served as subjects. They were all naïve with respect to this kind of experiment and were randomly allotted to the various experimental treatments.

Apparatus

The apparatus was a subject-controlled tachistoscope described fully by Naylor (1954, 1962). Briefly, it consisted of a light-proof box approximately 3 ft. in length. The stimulus lists could be inserted at one end of the box, and the subject could view them through an aperture at the other end. When he depressed a key a timer started and the list became visible. The releasing of the key after detection of the target item immediately cut off the illumination and stopped the timer.

Procedure

The subjects were seated comfortably in front of the apparatus and were instructed in its operation. The general nature of the task was explained to them.

Subjects in Experiment 1 were told that "nine of the 10 items in each list will be nonsense words," whilst those in Experiment 2 were told that "nine of the 10 items in each list will be meaningful words." Half the subjects in Experiment 1 were then told "to pick out the word which is meaningful" (positive instructions) and half "to pick out the word which is not a nonsense word" (negative instructions). Half the subjects in Experiment 2 were told "to pick out the word which is a nonsense word" (positive instructions) and half "to pick out the word which is not meaningful" (negative instructions).

Subjects were instructed to scan each list from the top, downwards, as quickly and accurately as possible. On detection of the target word, they were required to release their key and then spell out the word, letter by letter.

After familiarizing them with the apparatus, seven lists were presented to each subject with inter-presentation intervals of approximately 20 sec. No knowledge of results was allowed. The basic measure taken was the search time for each list. Errors of omission and commission were noted.

Results and Analysis

The measure employed in the analysis was mean time per item for each subject. This was calculated by (*a*) discarding all trials on which errors were made (errors were few, unrepresentatively long and tended to occur in the more difficult conditions) and (*b*) summing the time for all other trials and dividing this total by the total number of items which had to be scanned in order to detect the correctly reported targets. This was based on the assumption that the subjects had obeyed the instructions, i.e. that they had scanned each list from the top downwards (subjects did, in fact, take longer to find a critical item when it appeared lower in the list).

Experiment 1

Table I shows the mean time per item under each experimental condition. A summary of an analysis of variance of this data is shown in Table II. It is evident that both the main effects of target word and context were highly significant, i.e. it was easier to detect a HFW target than a LFW, and detection performance was facilitated when the context items were LANSs rather than HANSs. There was also a significant interaction between the type of target word and the type of instructions, i.e. the detection of a HFW target was facilitated with negative instructions and detection of a LFW target with positive instructions.

TABLE I

Mean Time per Item (sec.)

Experiment 1

Context	*HANS*		*LANS*	
Target	*HFW*	*LFW*	*HFW*	*LFW*
Instructions—				
Positive	0·82	0·78	0·59	0·65
Negative	0·62	0·88	0·40	0·77

TABLE II

Summary of Analysis of Variance

Experiment 1

Source	*S.S.*	*d.f.*	*M.S.*	*F*	*P*
Target (T)	0·218	1	0·218	8·72	<0·01
Context (C)	0·228	1	0·228	9·12	<0·01
Instructions (I) ..	0·015	1	0·015	—	—
T × I	0·183	1	0·183	7·32	<0·025
C × I	0·0003	1	0·0003	—	—
T × C	0·021	1	0·021	—	—
T × I × C	0·001	1	0·001	—	—
Error	0·609	24	0·025	—	—
Total	1·275	—	—	—	—

Experiment 2

A summary of means and analysis corresponding to Experiment 1 is presented in Tables III and IV for Experiment 2. Again, both the main effects of target word and context were significant, i.e. it was easier to detect a LANS target than a HANS target, and this process was facilitated when the context was made up of HFWs rather than LFWs.

TABLE III

MEAN TIME PER ITEM (sec.)

EXPERIMENT 2

Context	*HFW*		*LFW*	
Target	*HANS*	*LANS*	*HANS*	*LANS*
Instructions—				
Positive	0·63	0·59	1·19	0·85
Negative	0·89	0·58	1·12	0·80

TABLE IV

SUMMARY OF ANALYSIS OF VARIANCE

EXPERIMENT 2

Source	*S.S.*	*d.f.*	*M.S.*	*F*	*P*
Target (T)	0·500	1	0·500	10·417	<0·01
Context (C)	0·788	1	0·788	16·417	<0·001
Instructions (I)	0·008	1	0·008	—	—
T × I	0·029	1	0·029	—	—
C × I	0·063	1	0·063	—	—
T × C	0·047	1	0·047	—	—
T × I × C	0·043	1	0·043	—	—
Error	1·140	24	0·048	—	—
Total	2·618	—	—	—	—

DISCUSSION

From the results and analysis it can be seen that the predictions concerning the effects of the semantic disparity between a target and its context were verified. The detection of a HFW target was easier than that of a LFW target, and it was easier to search through a context of LANSs than HANSs. Conversely, it was easier to detect a LANS target than a HANS target, and this was further facilitated with a HFW context as opposed to a LFW context. Thus the two conditions in which the semantic disparity between a target and its context was greatest proved to be the most simple; whereas the two conditions in which semantic disparity was least, proved to be the most difficult.

The second main prediction—that positive instructions would lead to better search performance than negative instructions—was not supported by the results of either experiment. However, instructions must have been exerting some influence on performance as the data from Experiment 1 showed a significant interaction between type of instruction and target word frequency. Thus, if one is instructed to "pick out the meaningful word" it makes little difference whether this word is of high or low frequency. But, if one has to "pick out the word which is not a nonsense word" the task is made easier if the word is of high frequency.

One possible interpretation of this interaction lies in the fact that there is structural as well as semantic disparity between meaningful words and nonsense syllables. If a subject has to pick out the word which is not nonsense it is conceivable that his decision

to respond to the target is based on his testing of the word against his concept of a "nonsense" word (which has been built up during the course of the experiment). As a HFW is structurally more dissimilar from "nonsense" words than a LFW it might be expected that in this case the testing process is completed more rapidly.

In a similar way the absence of any effect of instructions in Experiment 2 can be accounted for as follows. Since subjects are basically unfamiliar with nonsense syllable targets their concept of these is necessarily formed throughout the experiment. Thus whether the label attached to them by the experimenter is "nonsense" or "not meaningful" the task from the subject's point of view remains essentially the same.

It is realized that this interpretation is somewhat *ad hoc* in nature and that doubtless there are others possible. However, it is clear that when a given target item is distinguished from its context in terms of *meaning*, the visual search process depends in some way on the type of instruction given.

References

Glaze, J. A. (1928). The association value of nonsense syllables. *J. genet. Psychol.*, **35**, 255–67.

Naylor, G. F. K. (1954). An approach to the study of dial reading. *Occup. Psychol.*, **28**, 90–8.

Naylor, G. F. K. (1962). Basic speed factors in perception. *Austral. J. Psychol.*, **14**, 101–12.

Neisser, U. (1963). Decision time without reaction time: experiments in visual scanning. *Amer. J. Psychol.*, **76**, 376–85.

Neisser, U. (1964*a*). Visual search. *Scient. Amer.*, **210**, 94–102.

Neisser, U. (1964*b*). Experiments in visual search and their theoretical implications. *Paper presented at the Psychonomics Society.*

Thorndike, E. L., and Lorge, I. (1944). *The Teacher's Wordbook of* 30,000 *Words.* New York.

Manuscript received 16th September, 1965.

THE EFFECT OF SET ON THE TWO-POINT TACTUAL THRESHOLD

BY

ANIMA SEN

From the Department of Psychology, University of Hull

This experiment studies the influence of set on the two-point tactual threshold. The two-point limen (critical stimulus) was determined along the mid-longitudinal line of both volar and dorsal surfaces of the right forearms of seven university students. Test-stimuli were selected at 3 mm. steps both up and down from each of the critical stimuli. Each of these test-stimuli was presented separately, the critical stimulus being interpolated 30 times in each test of the series. The proportions of two-point and one-point responses to the critical stimulus were then determined again in a control series. It was found that, as the test-stimulus deviated in the positive direction from the critical stimulus, the proportion of two-point responses to the critical stimulus increased, reached a maximum and then began to decline. A similar rise and fall in *one-point* responses were found in the opposite direction. The results were explained by Adaptation Level Theory.

Introduction

The influence of set derived from previous and contemporaneous experiences has been demonstrated in different fields of psychology by many investigators. Numerous studies (Gibson, 1929; Zangwill, 1937; Tresselt and Spragg, 1941; Bartlett, 1954; and others) indicate that the frequency and type of response to a particular stimulus may be markedly altered by the mental set or immediate aspect of a situation. Sequential dependencies on performance in a serial two-choice responding task also were demonstrated by Bruner and Wechsler (1958) and Bertelson (1961). The effect of set, or some predisposition, upon psychophysical data is also of great psychological interest. Cheesman and Mayne (1953) carried out an experiment indicating that the threshold for an olfactory stimulus was raised under the conditions of adaptation to a masking stimulus. Howarth and Treisman (1958, 1959) found that the auditory threshold was significantly lowered by a simultaneous visual stimulus which was itself well above threshold. It was shown by Dixon (1958) and Dixon and Haider (1961) that the visual thresholds for one eye varied as a function of the subliminal stimulation applied to the other. Helson (1959) reported that "past experience" influenced the succeeding judgements of lifted weights. But no reference could be found in the literature to any work on the influence of set or some other disposition in tactual perception, on two-point threshold in particular, though the interaction of time and space in cutaneous perception was determined by Wieland (1960).

Though the neurological mechanisms underlying two-point tactual discrimination cannot be exactly specified, very recently Zangwill and Gomulicki (1965) have shown that there is a correspondence between the cutaneous sensitivity as determined by the level of the two-point threshold, of a given body area, and the extent of its representation in the sensory cortex. They have found a striking resemblance between the Penfield cortical homunculi and their cutaneous sensitivity homunculus as translated from the numerical data. But the role of some non-anatomical factors on two-point limen were also demonstrated last century—by Volkman and Fechner as early as 1858, by Griesbach in 1895, and by Vannod in 1896 (all quoted by Boring, 1942). It was considered that the role of set in tactile perception of two-point threshold might be similarly significant. The aim of the present investigation was thus to determine the role of set developed in an experimental situation by the repeated presentation of two aesthesiometric points separated by an interval greater or smaller than the two-point threshold, upon the two-point tactual threshold.

Experimental Method

Subjects. Seven advanced students in psychology—five men and two women, aged 22–32, acted as subjects. Experiments with two other subjects were discontinued when it became apparent that their lack of interest was making them careless. The subjects had had experience of experiments of long duration. The purpose of the present investigation was not, of course, revealed to them.

Materials. (1) A Jastrow's two-point aesthesiometer; (2) a hand rest, and (3) a screen with a semi-circular aperture in the middle of the lower edge for passing the hand-rest through.

Procedure. The two-point limen (termed critical stimulus) was determined for each subject, by the method of limits, along the mid-longitudinal line of both volar and dorsal surfaces of the right forearm, half way down. The proportions of two-point and one-point responses to the critical stimulus thus determined was round about 50-50. For each subject, in each position, a range of stimuli (termed test-stimuli) was now selected on each side of the threshold, in 3 mm. steps, both up and down. In the effect-of-set situation each one of these test-stimuli was tested in a separate session; it was presented 90 times with the critical stimulus interpolated after every three consecutive presentations. Each time the subject was instructed to give his judgement by "two-point" when he felt two, and by "one-point" when he felt one. "Catch test" was given occasionally in order to ensure the subject's attentiveness. The possibility of a rhythmic expectancy had been excluded in a pilot investigation by the author (Chakraborty, 1963). The same procedure was followed with each test series. The required number of test-series in the upper and lower stimulus ranges, for both volar and dorsal surfaces, varied for each subject. The experiment was carried on systematically with progressively increasing test-stimuli (in the upper range) or progressively decreasing test-stimuli (in the lower range) for each subject as far as the test-series in which 85 per cent. or more of the responses to the critical stimulus were one-point ones (in the upper range) or two-point ones (in the lower range). In each session the selection of volar or dorsal surface was made by random assignment. Finally, there was a similar session for volar and dorsal surfaces separately, in which, instead of a single stimulus value, all of the stimuli were presented, in random order, so that the proportions of two-point and one-point responses to the critical stimulus could be re-determined without any effect of set. The percentages of two-point and one-point responses to the critical stimuli found in this control series were not always exactly the same values as determined by the limiting method; but the differences were statistically insignificant.

A session lasted about an hour, with brief rest-pauses after every 15 min. Each subject had only one session in a day, and from 34–40 sessions in all, spread over 8–12 weeks.

Analysis of Data and Results

The two-point limens of all the subjects for both volar and dorsal surfaces are given in Table I along with their percentages of two-point responses in the respective control series and in different test-series. The sequences of test-stimuli results have been centred at 0 (the critical stimulus) for each subject and the increase and decrease in test-stimulus values from the critical stimulus have been expressed in units of 3 mm. Two average curves for the results of volar and dorsal surfaces of all the subjects are shown in Figure 1. In drawing the average curves, the percentages of peak two-point responses (in the upper range) and those of lowest two-point responses (in the lower range) together with the percentages of two-point responses in the starting and end test-series were made to coincide for all the subjects.

Though the required number of test-series was different for each subject, the results in the table clearly show, at both test-sites in all subjects, the same systematic rise and fall in the percentage of two-point (in the upper range) or one-point (in the lower range) responses to the critical stimulus. In every case, the difference between the proportion of two-point responses to the critical stimulus in each test-series and the proportion in the corresponding control series was separately tested by the *t*-test for statistical significance and the significantly increased or decreased percentages are shown by asterisks in the table. In the upper ranges of stimuli, the proportion of two-point responses to the critical stimulus increased in consecutive test-series, step by step for several steps, to reach a peak that was statistically highly significant (with the exception of subject C—dorsal and subject D, volar). This marked rise in two-point proportions was followed, in every case, by a steady decline, which was paradoxical in view of the continued increase in the distance between the stimulating points. In this decline there was a neutral or "indifferent range," in which, for several stimulus steps (different for different subjects), the two-point proportion was not significantly higher or lower than that in the control series. Then, with still greater stimulus-distances, the two-point proportions fell far below the proportion in the control series, the differences being highly

TABLE I

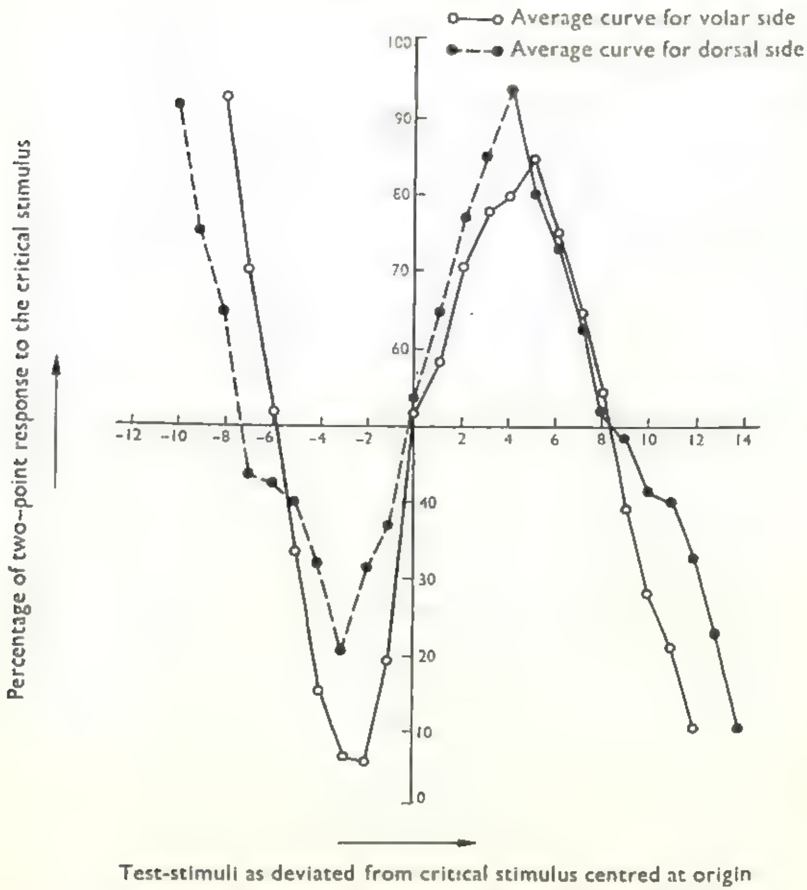
CRITICAL STIMULI AND THE PERCENTAGE OF TWO-POINT RESPONSES TO THEM IN DIFFERENT TEST SERIES AND CONTROL SERIES

Subject	Side †	Percentage of two-point response in different test-series of the lower range									Critical stimulus value (in mm.)	Percentage of two-point response in control series	Percentage of two-point response in different test-series of the upper range													
		−9	−8	−7	−6	−5	−4	−3	−2	−1			+1	+2	+3	+4	+5	+6	+7	+8	+9	+10	+11	+12	+13	+14
A	V						** 90·00	66·67	33·33	* 16·67	21·0	46·67	53·33	* 73·33	* 73·33	** 80·00	** 83·33	* 76·67	* 73·33	63·33	23·33	** 13·33				
	D	* 86·67	70·00	63·33	43·33	43·33	40·00	33·33	** 23·33	36·67	29·5	56·67	73·33	** 96·67	** 100·00	** 100·00	** 93·33	80·00	* 30·00	* 26·67	** 13·33					
B	V			** 90·00	* 76·67	36·67	** 3·33	** 0·00	** 3·33	** 13·33	33·5	50·00	53·33	63·33	* 80·00	** 90·00	** 83·33	63·33	43·33	33·33	* 23·33	* 20·00	** 13·33			
	D				** 96·67	73·33	66·67	40·00	** 23·33	* 33·33	29·0	63·33	76·67	* 86·67	** 96·67	** 100·00	** 96·67	80·00	* 90·00	56·67	60·00	43·33	40·00	** 16·67	** 6·67	
C	V			** 90·00	* 76·67	40·00	* 23·33	** 3·33	** 0·00	** 3·33	30·5	50·00	66·67	* 80·00	** 83·33	* 76·67	66·67	66·67	60·00	43·33	33·33	26·67	** 3·33			
	D				** 86·67	* 80·00	30·00	26·67	33·33	36·67	27·0	50·00	70·00	63·33	66·67	46·67	* 20·00	** 10·00								
D	V				** 93·33	66·67	* 26·67	** 6·67	** 10·00	33·33	19·5	53·33	53·33	70·00	66·67	60·00	53·33	53·33	43·33	* 26·67	** 16·67	** 10·00				
	D				** 100·00	70·00	70·00	33·33	** 6·67	43·33	27·5	46·67	53·33	70·00	* 76·67	** 86·67	* 76·67	70·00	56·67	50·00	43·33	43·33	40·00	33·33	26·67	** 13·33
E	V				** 96·67	73·33	50·00	30·00	** 16·67	33·33	27·0	50·00	56·67	73·33	** 90·00	** 93·33	** 96·67	** 86·67	* 76·67	66·67	66·67	40·00	36·67	** 13·33		
	D					** 96·67	* 83·33	63·33	43·33	* 23·33	27·5	53·33	53·33	** 93·33	73·33	56·67	50·00	46·67	30·00	* 23·33	** 6·67					
F	V				** 100·00	53·33	** 20·00	** 0·00	** 6·67	** 16·67	30·0	60·00	56·67	70·00	70·00	* 83·33	53·33	50·00	46·67	40·00	** 23·33	** 13·33				
	D			** 86·67	60·00	43·33	30·00	** 13·33	* 26·67	43·33	28·0	53·33	63·33	73·33	76·67	* 83·33	70·00	56·67	44·67	36·67	** 20·00	** 6·67				
G	V						** 90·00	60·00	** 13·33	* 20·00	15·5	50·00	66·67	73·33	** 86·67	* 80·00	60·00	46·67	33·33	* 20·00	** 16·67	** 6·67				
	D				** 90·00	73·33	43·33	30·00	33·33	40·00	21·0	53·33	63·33	76·67	76·67	** 100·00	* 80·00	66·67	40·00	33·33	** 20·00	** 6·67				

† V = Volar
D = Dorsal

* $p < 0{\cdot}05$.
** $p < 0{\cdot}01$.

FIGURE 1



Two average curves for Volar and Dorsal sides.

significant statistically. Although the number of test-series in every lower range of stimuli was smaller than that in the corresponding upper range, a comparable sequence of fall and rise in two-point responses was observed.

Though the two-point thresholds differed from site to site and subject to subject, the two average curves show the same trend which was observed in individual cases—the rise and fall (in the upper range) and fall and rise (in the lower range) in two-point responses.

Discussion

The discrepancies between the subjects' responses in different test-series, and those in the respective control series, clearly showed that the subjects developed a set or predisposition to a two-point response (in the upper ranges) or a one-point response (in the lower ranges) to the critical stimulus. If there were no set, the numbers of two-point (or one-point) responses to the critical stimuli in different test-series would closely approach those in the respective control series. When, without the subject's knowledge, the critical stimulus was applied to him at every fourth presentation, he displayed a tendency from his usual expectation to respond to the higher (or lower) test-stimulus. The results showed that as the test-stimulus became greater the proportion of two-point responses to the critical stimulus rose to a peak then fell, through a neutral or indifferent range, to 15 per cent. or lower. Similarly, in the lower ranges, as the test-stimulus became progressively smaller, the proportion of one-point responses rose to a peak and then fell, through an indifferent stage, to less than 15 per cent. The results thus showed that the set operated within an optimal range, and after that it had no effect for several stimulus-steps, and then it exerted its influence in the opposite direction. The present results

had a feature in common with the results obtained by Cheesman and Mayne (1953) who found that a masking stimulus raised the olfactory threshold for a particular stimulus. The shift from a positive to a negative effect of set observed in the present study, as the test-stimulus was progressively increased (in the upper range) or decreased (in the lower range), reminds one of Jarvik's (1951) report of a change from a positive to a negative reinforcement in a probability series.

The present results can be explained by the theory of Adaptation Level (Helson, 1947) with wider applicability. The notion of Adaptation Level has not been restricted to the effects of prolonged adaptation to some constant stimulation resulting in greatly reduced capacity for response. The concept of AL has also been extended to explain some factors like "past experience," expectancy levels, etc. An AL is present for every moment of stimulation, changing in time and with varying conditions of stimulation, and it is a function of all the stimuli acting upon the organism at any given moment as well as in the past. For every excitation-response configuration there is assumed a stimulus which represents the pooled effect of all the stimuli and to this the organism is said to be adapted or attuned. Stimuli near this value bring forth some neutral responses. In the present experiment also, the subjects were attuned to the test-stimulus of different test series before responding to the critical stimulus. But when the test stimulus was nearer to the critical stimulus, it could not exert its influence on the response to the critical stimulus and the nature of responses to the latter became "neutral," i.e. the percentages of two-point response did not differ significantly from those in the control series. But as the test-stimuli increasingly deviated from the critical stimulus on both sides, they exerted positive influence on the response to the critical stimulus. The test-stimuli of the upper ranges brought mostly two-point responses to the critical stimulus. Similarly, the smaller stimuli of the lower range established negative gradients bringing mostly one-point responses to the same. This positive enhancement in two-point or one-point responses continued for some time as the subjects developed a set at a given level (in the test series of upper or lower range) from previous stimulations. The result showed clearly that the two-point responses to the critical stimulus increased step by step for a few steps in the upper range test-stimuli. Similarly, one-point responses increased in the lower range test-stimuli through a comparatively less number of stimulus steps. But the set operated its influence within some optimal range and again a neutral range appeared after which the negative influence of set became apparent. The negative effect of set can also be explained by the negative effects in AL theory. When a stimulus is far above the stimulus range, the AL is raised so high that it sensitizes to negative qualities while the opposite is the case with a stimulus far below the range. In the present experiment also it is seen that when a test-stimulus was far above the critical stimulus, the majority of the responses to the latter was of negative kind, i.e. one-point ones. Similarly, a test-stimulus far below the critical stimulus brought mostly two-point responses to the latter.

Thus the present study showed that the percentages of two-point or one-point response to a two-point threshold stimulus are greatly altered by the set developed from the previous higher or lower stimulations. So the response to the two-point threshold stimulus is not always an absolute judgement. A "frame of reference" or background may markedly alter the nature of the response to the threshold stimulus. The value of two-point thresholds differed from site to site and subject to subject and the sequences of appearance of peak and lowest two-point responses together with the turning-points showing the negative effect were also different. It would be interesting to establish mathematically a relationship between the size of the two-point threshold and the optimum range showing the effect of set using different body parts, and, if possible, with more subjects.

This research was carried out in the Psychology Department of Calcutta University under the guidance of Prof. K. C. Mookerjee in partial fulfilment of the requirements for the degree of D.Phil. I am very grateful to Dr. A. M. Clarke, Prof. A. D. B. Clarke, Dr. G. H. Wright and Mr. A. K. Sen, for their criticisms and many helpful suggestions.

References

Bartlett, F. C. (1954). *Remembering*. London: Cambridge University Press.

Bertelson, P. (1961). Sequential redundancy and speed in a serial two-choice responding task. *Quart. J. exp. Psychol.*, **13**, 90–102.

BORING, E. G. (1942). *Sensation and Perception in the History of Experimental Psychology*. New York: Appleton Century Croft.

BRUNER, J. S., and WECHSLER, H. (1958). Sequential probability as a determinant of perceptual closure. *Amer. J. Psychol.*, **71**, 604–6.

CHAKRABORTY, A. (1963). A comparative study of the effects of the impression stimuli presented in systematic and haphazard orders on two-point tactual threshold stimulus. *Ind. J. Psychol.*, **38**, 193–6.

CHEESMAN, G. H., and MAYNE, S. (1953). The influence of adaptation on absolute threshold measurements for olfactory stimuli. *Quart. J. exp. Psychol.*, **5**, 22–30.

DIXON, N. F. (1958). Apparent changes in the visual threshold as a function of subliminal stimulation. *Quart. J. exp. Psychol.*, **10**, 211–9.

DIXON, N. F., and HAIDER, M. (1961). Changes in the visual threshold as a function of subception: a preliminary report. *Quart. J. exp. Psychol.*, **13**, 229–35.

GIBSON, J. J. (1929). The reproduction of visually perceived forms. *J. exp. Psychol.*, **11**, 1–39.

HELSON, H. (1947). Adaptation level as frame of reference for prediction of psychophysical data. *Amer. J. Psychol.*, **60**, 1–29.

HELSON, H. (1959). Adaptation level theory. In KOCH, S. (Ed.), *Psychology: A Study of a Science*, 1. New York: McGraw Hill.

HOWARTH, C. I., and TREISMAN, M. (1958). The effect of warning interval on the electric phosphene and auditory thresholds. *Quart. J. exp. Psychol.*, **10**, 130–41.

JARVIK, M. E. (1951). Probability learning and a negative recency effect in the serial anticipation of alternative symbols. *J. exp. Psychol.*, **41**, 291–7.

TREISMAN, M., and HOWARTH, C. I. (1959). Changes in threshold level produced by a signal preceding or following the threshold stimulus. *Quart. J. exp. Psychol.*, **11**, 129–42.

TRESSELT, M. E., and SPRAGG, S. D. S. (1941). Changes occurring in the serial reproduction of verbally perceived materials. *J. genet. Psychol.*, **58**, 255–64.

WIELAND, B. A. (1960). The interaction of space and time in cutaneous perception. *Amer. J. Psychol.*, **73**, 248–55.

ZANGWILL, O. L. (1937). A study of the significance of attitude in recognition. *Brit. J. Psychol.*, **28**, 12–7.

ZANGWILL, O. L., and GOMULICKI, B. R. (1965). The Penfield "Homunculus" and the two-point threshold. Unpublished work.

Manuscript received 5th November, 1965

"ADAPTATION AND REPULSION IN THE FIGURAL AFTER-EFFECT" AND THE PSYCHOPHYSICAL THEORY

BY

M. M. TAYLOR

From Defence Research Medical Laboratories, Toronto, Ontario

Data recently presented by Wilson (1965) seem to demonstrate the separate effects of adaptation and of after-effect repulsion during and following continued observation of a curved line. Inasmuch as the experiment was performed without apparent reference to the psychophysical theory of figural after-effects (Taylor, 1962), it is interesting to note that the results on adaptation agree qualitatively with one of the major presuppositions of the theory, and the results on repulsion agree quantitatively with its predictions.

Neutralization and Displacement

One of the major presuppositions of the psychophysical theory of figural after-effects is that there are at least two separable phenomena, "neutralization and displacement. These two effects are phenomenologically very similar, and often have been considered the same. Neutralization may be described as a regression of the percept towards a 'norm' with continued inspection of a constant stimulus. Displacement is a shift of a percept toward or away from another percept when compared to the percept produced by the same stimulus in isolation. Both neutralization and displacement may involve changes in any perceptual dimension. Neither is restricted to geometric location, though most studies of figural after-effects have used geometric location as a criterion variable" (Taylor, 1962, p. 248). The quantitative part of the theory dealt only with the displacement effect, considering the neutralization (Wilson's (1965) and Gibson's (1933) "adaptation") to be an independent and unrelated effect.

Wilson (1965) has demonstrated that for curvilinearity both adaptation (neutralization) and repulsion (displacement) take place with essentially continuous passive observation of a curved inspection line. Lines of curvature in the same direction as the inspection line are seen as less curved after than before inspection, but added to this tendency is a displacement such that the reduction in curvature is greater for lines less curved than the inspection line and less for lines more curved than the inspection. This result is in exact accord with the presuppositions of the psychophysical theory, and agrees with experimental results for the after-effect of seen motion (Carlson, 1962; Rapoport, 1964; Taylor, 1963*a*, 1963*b*).

It was argued in Taylor's papers, and by Taylor and Ross (1964) that the motion after-effect is an example of the figural after-effects. Wilson's results support that contention by displaying the separate adaptation and repulsion phenomena. If the motion and the curvature after-effects are simply members of the larger class of figural after-effects, however, Wilson's interpretation of adaptation and repulsion as derived from the same mechanism seems incorrect. Particular differences between adaptation and repulsion in the motion after-effect include the time-scale (the after-effect reaches its maximum magnitude within 1 min. of inspection, whereas the neutralization shows no signs of reaching any asymptote within 10 min.), and the effect of varying inspection speed (neutralization is in proportion faster the slower the inspection speed, while the after-effect is greatest at a medium speed, decreasing at faster and slower inspection speeds).

Quantitative Prediction of Displacement

One presumption of the psychophysical theory is thus supported by Wilson's results. From more definite assumptions, an equation was derived from the theory (Taylor, 1962, p. 265). This equation related the amount of after-effect to the difference between inspection and test figures on some measurement scale, and to variations in the observer's acuity for differences on that scale. The equation is:

$$E = \frac{hM/R}{1 + (kM/d)^2} e^{-t/(b\sqrt{I})}$$

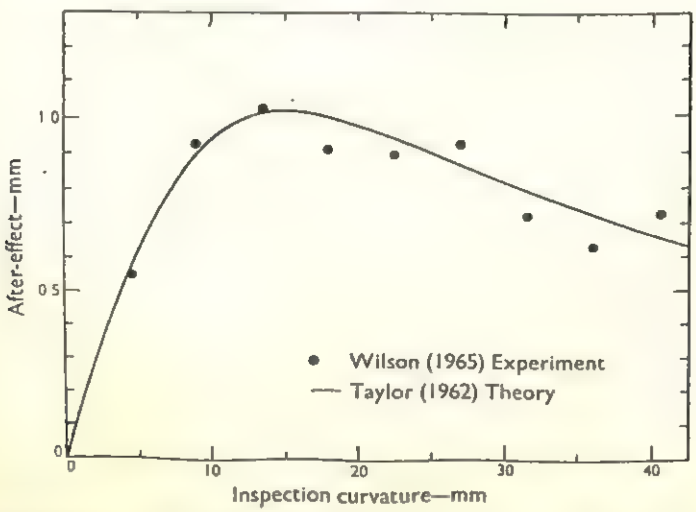
where E is the measured after-effect, M the separation on the same scale between inspection

and test figures, d the jnd for variations on the same scale at the position of the test figure, R the ratio of jnds for the test and the inspection figures, t the time since the end of the inspection period, and I the length of the inspection period. The other symbols, h, k, and b are fitting parameters. The parameter b affects only the time scale of the decay, and not the shape of the distance paradox function. To fit a distance paradox function, the exponential part of the expression takes some constant value less than unity. In a particular experiment, the value may be guessed by analogy with results in other experiments.

To attempt a fit to Wilson's results (his Fig. 4), the value of the decay time constant was taken as 8·3 sec. This value was chosen as representative of the values displayed by Taylor (1962, Fig. 2) for analogous studies with 15 sec. inspection period. With a delay of 4 sec. between the end of the inspection period and the start of the test period, the after-effect would be expected to decay to 0·62 of its original value. By the end of the test period, 1 sec. later, it should have decayed to 0·57 of its original value. These figures are probably underestimates of the measured after-effect, and overestimates of the decay, because the technique of repeated measurement seems conducive to carryover from one inspection period to the next, with a consequent increase in the decay time constant. (Wilson tested for an increase in the after-effect by comparing the first half with the second half of the run. This test would not be expected to detect the carryover of the effect, which should be near asymptote by about the third response.)

In the fit of the equation to Wilson's data, the value of d is unknown. Accordingly, (k/d) was treated as a single parameter, and in the absence of contrary information R was assumed to be unity. The decay was assumed to be to 0·60 of the initial value. The fit obtained was computed with parameter values h = 0·226, k/d = 0·067. In previous applications of the equation to experiments of this type (Taylor, 1962, 1963c), h has usually been found to be about 0·2, in reasonable agreement with the present value of 0·226. k has ordinarily been found to be 0·03, which with the present fit implies that d, the jnd for curvilinearity, is 0·45 mm. In the 19 cm. length of curve used by Wilson, this value seems of the right order of magnitude, especially since Wilson using the method of constants selected as stimuli six comparison lines differing by 0·24 mm. The actual value of d could be computed, given the 25 per cent. and 75 per cent. points (not published) on Wilson's probit lines.

FIGURE 1



After-effect curvature of a straight test line as a function of the curvature of the inspection line. Data from Wilson (1965), curve from equation given by Taylor (1962). Parameter values: $(h.e.^{-t/(b\sqrt{I})}) = 0{\cdot}136$; $(k/d) = 0{\cdot}067$.

In Figure 1 are shown Wilson's experimental points from his Figure 4, combined for left and right curvature, together with the curve derived from the psychophysical theory of figural after-effects. The scatter of points about the curve suggests that the fit is

within the experimental error. As well as can be shown from the data Wilson presents, the two parameter values needed for the fit agree with those used to fit the same equation to results of other experiments (Taylor, 1962, 1963*c*).

Wilson's results both for adaptation and for repulsion appear to give support to the psychophysical theory of figural after-effects.

This is Defence Research Medical Laboratories Research Paper No. 611.

References

Carlson, V. R. (1962). Adaptation in the perception of visual velocity. *J. exp. Psychol.*, **64**, 192–7.

Gibson, J. J. (1933). Adaptation, after-effect, and contrast in the perception of curved lines. *J. exp. Psychol.*, **16**, 1–31.

Rapoport, J. (1964). Adaptation in the perception of rotary motion. *J. exp. Psychol.*, **67**, 263–7.

Taylor, M. M. (1962). Figural after-effects: a psychophysical theory of the displacement effect. *Canad. J. Psychol.*, **16**, 247–77.

Taylor, M. M. (1963*a*). Tracking the decay of the after-effect of seen rotary movement. *Percep. mot. Skills*, **16**, 119–29.

Taylor, M. M. (1963*b*). Tracking the neutralization of seen rotary movement. *Percep. mot. Skills*, **16**, 513–9.

Taylor, M. M. (1963*c*). Numerical prediction of a simple figural after-effect as a function of the contrast of the inspection figure. *Psychol. Rev.*, **70**, 357–60.

Taylor, M. M., and Ross, P. L. (1964). Tracking rotary motion after-effect with different illuminations of inspection and test fields. *Percep. mot. Skills*, **18**, 885–8.

Wilson, J. (1965). Adaptation and repulsion in the figural after-effect. *Quart. J. exp. Psychol.*, **17**, 1–13.

Manuscript received 20th September, 1965

COMMENTS ON DR. M. M. TAYLOR'S NOTE

BY

J. WILSON

From the Department of Psychology, University of Bristol

Taylor's psychophysical theory "attempts to explain only the displacement effect" (Taylor, 1962), and therefore has no more relevance to the relation between adaptation and repulsion, which was the subject of my paper, than the other repulsion (displacement)* theories which I used as examples. However, Taylor's theory differs from the other repulsion theories in providing a connection between the figural after-effect and changes in discriminability. In this respect it has had some success with a variety of experimental situations, and it is gratifying to learn that when applied to one of my experiments it gives a theoretical curve which fits well, both in form and in approximate magnitude.

However, Taylor claims stronger support from my data than is justified. First, he maintains that my results on adaptation agree qualitatively with one of the "major presuppositions" of his theory. But this major presupposition amounts only to a recognition of the existence of adaptation and a claim that it is separate from the repulsion. My experiments, among others, show that both repulsion and adaptation play a part in the figural after-effect, and that a gradation of situations exists: between those situations in which repulsion predominates and those in which adaptation predominates there is a full range of intermediate situations. My experiments do *not* show that adaptation and repulsion are separate phenomena, and I know of no experiment which does. In any case, it seems to me that the "major presupposition" has nothing to do with Taylor's psychophysical theory itself; the presupposition serves only to exclude the adaptation effects so that the theory, dealing only with repulsion effects, can be safely applied.

I have always thought, as Taylor does, that the motion and curvature after-effects are simply members of the larger class of figural after-effects; but this does not in any way affect my argument that adaptation and repulsion may be due to the same mechanisms. To show that this argument is erroneous, Taylor cites a difference in time-scale between adaptation and repulsion in motion after-effects; but the necessity for a difference of this kind is an essential part of my argument and is explicitly stated in my paper (see p. 10, last paragraph). Moreover, it is clear that such differences in time-scale exist even within repulsion effects only, since it is possible to change the rate of decay of an after-effect by merely changing the inspection time. It is difficult to understand Taylor's point about the effect of varying inspection speed, since an "in proportion" variation of time rate cannot strictly be compared with a magnitude of effect. Thus I can see no reason for abandoning my argument that adaptation and repulsion may be produced by the same mechanisms. Moreover, it seems to me that my argument does not involve the psychophysical theory as such in any particular difficulty.

Taylor's second claim is that my results on repulsion agree quantitatively with the predictions of his theory. To some extent he has justified this claim; but the data which he uses, from my Figure 4, are the results of an experiment using curved inspection lines, an experiment in which therefore some part of the after-effect is due to adaptation. Had I eliminated adaptation effects by using a straight inspection line and curved test lines, the magnitude of the after-effect would have been smaller by about a third (as can be seen from the other experiments reported in my paper), and Taylor's formula would then have required different values for the fitting parameters. However, it is clear that Taylor's theoretical curve provides a very good fit to the data.

Reference

TAYLOR, M. M. (1962). Figural after-effects: a psychophysical theory of the displacement effect. *Canad. J. Psychol.*, **16**, 247–77.

Manuscript received 6th October, 1965

* Taylor's term displacement is not identical in meaning with my term repulsion, but for present purposes I have followed him in identifying them.

A NOTE ON DAVIS'S REFUTATION OF THE EXPECTANCY HYPOTHESIS

BY

JOHN ANNETT

From The Department of Psychology, The University, Hull

Davis (1965) presents an experimental situation designed to distinguish between the intermittency hypothesis and the expectancy hypothesis in accounting for abnormally long reaction times when the interval between the stimulus and a prior warning signal is very short. He argues that if refractoriness is due to central processing of the first stimulus, or warning signal, there should be no delays to reaction time if this first stimulus is eliminated. To provide a fair test of the expectancy hypothesis it is necessary to have an event specifying the beginning of the waiting period and Davis attempts to meet this condition by having the waiting period begin, not with an externally applied stimulus but with a key-pressing response spontaneously emitted by the subject himself. Then . . . "If the distribution of intervals is kept exactly the same as in the more usual situation where two signals are represented, any difference in the pattern of the reaction time between the two situations must be attributed to the effect of the event which commences the interval rather than the distribution of intervals used."

Davis's data show no refractory effect for the spontaneous response situation thus implying support for the intermittency hypothesis. The argument on which his test is based is, however, open to question. The difficulty lies in the acceptance of the subject's response as the event defining the beginning of the waiting period. Experimentally it is only possible to specify the precise time at which the contacts on the subject's key are closed. This is a physical event and not an event in the subject's nervous system—indeed we can only guess at the prior chain of events inside the subject and we certainly cannot define their beginning with any degree of precision. Moreover we must be sure that the event taken by the experimenter to mark the beginning of the waiting period (the closing of the key contacts) is the same as that taken by the subject. One must presume that Davis was careful to eliminate any stimulus event (a click or a sudden change in the key resistance) associated with the closing of the key, otherwise he would have had a stimulus marking the beginning of the waiting period which, it would seem, must be specifically excluded to test this hypothesis. Having taken pains to conceal from the subject the event defined by the experimenter as the beginning of the waiting period how can we know that this coincides with the subject's definition of the beginning of the waiting period? Indeed, one might ask how can the subject himself know under these circumstances?

Thus whether or not the *purely efferent* aspects of a response can be taken as an event capable of marking, for the subject, the commencement of a waiting period we are still left with the problem that the experimenter may not be able to know when this "event" occurs, or more precisely, what its relationship is with the closing of the response switch. If the subject is persuaded to make a rapid tonic response the experimenter might reasonably assume that there is a short and roughly constant time between the efferent activity and the closing of the switch. But in this situation it would be virtually impossible to avoid accompanying proprioceptive stimuli. A slower controlled response on the other hand, whilst providing no clearly differentiated stimulus event, also leaves the response event vaguely defined. Since we are dealing with intervals of 50 to 500 millisec. very little time uncertainty can be tolerated if the expectancy theory is to be given a fair test. The matter of eliminating any stimulus event is, therefore, much more diffcult than Davis would have us believe.

Thus we have a dilemma. If there is no specifiable event (either externally induced or feedback from a response) coinciding with the experimenter's definition of the beginning of the waiting period, expectancy theory has no case to answer. The subject is presumed to be performing under simple Reaction Time conditions without a specified warning period. If, on the other hand, there is a stimulus associated with the subject's spontaneous response, even if only proprioceptive, the intermittency theorist must explain why this stimulus does not cause refractoriness whilst a light flash or buzz does.

The dilemma seems almost inescapable but a possible way out would be to assume that a stimulus arising immediately from a spontaneous response is so redundant that it does not occupy the central channel and thus causes no refractoriness. In this case there should be no difference between a silent and a noisy response key for a highly practised subject. Whatever the respective merits of the intermittency and expectancy theories Davis's experiment, as it stands, does not distinguish between them.

REFERENCE

DAVIS, R. (1965). Expectancy and intermittency. *Quart. J. exp Psychol.*, **17**, 75–8.

Manuscript received 6th November, 1965

A REPLY TO DR. JOHN ANNETT

BY

R. DAVIS

From the Department of Psychology, The University of Reading

If the argument is about specifying the temporal uncertainty of the critical signal in relation to some neural event, then the difficulties Annett mentions arise not only in the spontaneous response situation, but also in all other situations previously used. Thus, in the comparison experiment where two signals were presented in sequence and a response required to the second critical signal, there would be a similar problem of identifying the "neural event" corresponding to the first signal. However, this has apparently not troubled those who have sought to explain delays in responding to the second of two signals in terms of expectancy.

It is true that there would be a difference between the two-signal situation and the spontaneous response situation if the time to the critical signal were measured from the hypothetical neural event rather than from the moment of occurrence of the first signal or the moment of occurrence of the spontaneous response. In the two-signal situation, the neural event would occur after the first signal so that the time from it to the critical signal would be *shorter* than the measured time between signals. In the spontaneous response case the neural event would occur before the spontaneous response so that the time from it to the critical signal would be *longer* than the measured intervals in my experiment. Karlin (1965) has made a rather similar point in his comments.

However, the consequences of this are not at all in favour of an expectancy theory. The effects of temporal uncertainty are known to increase with an increase in the range of intervals used and, therefore, there should be a greater effect due to temporal uncertainty in the situation with the longer range of intervals (the spontaneous response situation) than in the situation with the shorter range of intervals (the two-signal situation).

The results were in precisely the opposite direction. Hence the attempt to make use of a hypothetical neural event as the point from which intervals should be measured seems to me to make the expectancy case even less plausible.

REFERENCE

KARLIN, L. (1965). Comments on "Expectancy and intermittency." *Quart. J. exp. Psychol.*, **17**, 351.

Manuscript received 4th December, 1965

BOOK REVIEWS

The Nature of Psychology—A Selection of Papers, Essays and Other Writings by Kenneth J. W. Craik. Edited by Stephen L. Sherwood. London: Cambridge University Press. 1966. Pp. xx + 184. 30s.

To sort, select and publish material from the unfinished writings of Kenneth Craik is an editorial task well worth doing and in this instance well done. K. J. W. Craik, as almost all the readers of this *Journal* will know, was a psychologist of great promise who died in a motor accident in 1945 at the age of 31. A classics scholar who read philosophy under Kemp Smith at Edinburgh, Craik was a natural experimentalist and mechanic who in the intellectual ferment of the last war was able to capture for himself and for his subject the interest and support of physiologists and physical scientists. Hence, in spite of—or perhaps because of—his early tragic death, he became in a sense a symbol which marked the return in this country of psychology to the scientific fold.

To the reviewer who knew Craik first socially and only posthumously as a scientist, these papers are curiously provocative. As scientific papers they, not surprisingly, lack the polish and incisiveness of Craik's published work. As evidence of Craik's methods of work, the absence of biographical detail and explanatory comment is a serious defect. How much in fact did Craik amend and polish? One would have welcomed the reproduction of a page of Craik's "often difficult handwriting". A portrait would also have been more than welcome. As the editor says, the possibility of observing the birth, growth and mutation of Craik's ideas is fascinating. Unfortunately, the success the editor had in doing this is not conveyed to the reader. One of the few editorial comments is to the effect that Craik's writing contained "ten recorded presentiments" of the manner of his death. To the reviewer this seems bizarre: certainly the example given is in the worst traditions of research on precognition.

The format of the book is excellent and the price not excessive. Perhaps therefore the reviewer has been unfair in asking for more. In other words, the aim and execution of presenting the text of Craik's unpublished work is well carried out and the absence of biographical detail is in part made good by the reprinting of the evocative obituary which Sir Frederic Bartlett wrote for the British Journal of Psychology. Let us hope that it will not be long before the official reports on which Craik's reputation partly rests and which are not included in the present volume can be made more widely available.

A. Elithorn.

Mach Bands: Quantitative Studies on Neural Networks in the Retina. By Floyd Ratliff. San Francisco, London and Amsterdam: Holden-Day. 1965. Pp. xii + 365. $15.25.

Ernst Mach has good claims to be regarded as a major prophet of modern science. He is widely recognized as Einstein's John-the-Baptist, and it cannot be dismissed as pure chance that cosmologists lecture us about Mach's Principle while aeroplanes exceeding Mach One thunder at us from the sky, and this book is written on Mach Bands. Any psychologist who has dipped into *The Analysis of Sensation* knows that this prophet speaks with a very modern voice, and he will open Floyd Ratliff's book expecting further inspiration. He will, however, find the tone scholarly and academic, rather than prophetic. *Mach Bands* demands sober and lengthy study, rather than brief perusal for prophetic insight.

The book is divided into two parts: the first (243 pp.) is made up of five chapters by Ratliff; the second (80 pp.) the translation, also by Ratliff, of six papers by Mach that are not readily available to the average reader. These are followed by a useful bibliography and index. The first five chapters start with an account of Mach's life and, though brief, I found this filled many gaps in my sketchy knowledge of Mach's background. It is followed by a description of the ways in which a linear change of luminance can be produced, and a survey of some of the recent psychophysical investigations of such Mach-band producing displays.

The third chapter deals with various mathematical approaches to Mach bands. These are usefully classified as linear or non-linear, recurrent or non-recurrent, but I finished this chapter with a sense of disappointment. A good mathematical representation should succinctly unify all the experimental results on the phenomenon, and one may reasonably hope for insight into what will happen in a new situation. It seems, however, that the

mathematical representation is either concise and inaccurate, or precise and hopelessly unwieldy; in this respect it is, unfortunately, a typical product of the engineering approach to biological problems.

The fourth chapter is on the functional significance of inhibitory interaction. The main points Ratliff makes are that lateral inhibition accentuates the borders and contours of objects, thereby aiding their recognition, and that it undoes the blurring introduced by imperfect optics. There is also a detailed comparison of the contrast mechanism in Limulus and human, and a discussion of some of the more complex image processing that occurs in the retina and cortex. Though thorough and accurate, I feel that something is missing from this chapter, for there is more to be said about Mach bands than a pure physiologist will ever say. If retinal mechanisms abstract contours, it is tautologous to point out that contours are used in recognition: the brain can only use what the eye passes on to it, and there is thus no alternative to the utilization of contours. What urgently requires explanation is "Why use contours?" "What is the survival value of contour abstraction?"

The final chapter is on "Appearance and Reality." It is largely made up of historical accounts of the interesting errors that have arisen through assuming that subjectively observable Mach bands correspond to physical maxim and minim of intensity. These bands are amongst the most vivid of illusions perhaps because they originate very early in the visual system—but I doubt if their study contributes in an important way to the philosophical problem. On the other hand what is omitted from the earlier chapter does seem philosophically relevant, and relevant to one of the ideas from Mach's own fertile mind. He held that scientific hypotheses, as well as many everyday concepts, were essentially a means of achieving "economy of thought." It is at least worth entertaining the notion that lateral inhibition—the mechanism responsible for Mach bands—earns its survival value by making an early step towards such Machian economy.

In summary the main part of this book is a scholarly and professional treatment of a topic of considerable current interest; but I am not sure how much this very professionalism would have appealed to Mach, a man who described himself as "merely an unprejudiced rambler, endowed with original ideas." Fortunately the preliminary chapter gives us an impression of the scope of Mach's work, and the translated papers an idea of its quality; his unprejudiced rambling seems to lead one more directly to the core of a problem than the disciplined march of modern scientific professionalism. HORACE B. BARLOW.

Speech Analysis, Synthesis and Perception. By J. L. Flanagan. Berlin, Heidelberg and New York: Springer-Verlag. (Distributors in U.S.A. and Canada: Academic Press, N.Y.). 1965. Pp. viii + 317. DM. 58.

Since the end of the war, the study of hearing has changed very considerably. Some of these changes are due to studies of the ear itself, showing that one cannot regard it as a set of piano wires resonating each in isolation to a particular frequency, but rather as a complex hydrodynamic system through which waves travel from end to end. Other changes have resulted from psychophysical studies of complex sounds rather than pure tones, and from the advances of telephone engineers in constructing systems which will produce speech-like sounds. Relatively few psychological textbooks take note of these changes, and it is not easy to find a single brief reference to which to refer enquirers.

Most of these more recent developments are mentioned in this book, and for those who are prepared to spend the effort it may well be a useful survey of the change in knowledge of speech in the past 15 years or so. The plan of the book is that it deals first with the way in which different speech sounds are articulated, concentrating primarily upon an analysis in acoustic terms, by writing equations and analogue circuits for the vocal tract. It then deals with the mechanism of the ear, rather in the same way, and with the various alternative ways of analysing the speech sounds acoustically. The different possible ways of constructing machines to synthesize speech are then discussed, and there is some consideration of the limits of the human listener in identifying speech-like sounds: this includes a survey of the Haskins laboratory work on the various acoustic cues which give rise to perception of particular sounds. Lastly, there is a discussion of complete telephone systems which will analyse the voice, break it down into relatively few signals for transmission, and then reassemble it at the far end of the line.

As might be suspected from this outline, the book is not primarily intended for psychologists, and one would guess that the intended audience is probably one of fairly sophisticated engineers and acousticians who happen to be rather ignorant of speech. Thus the mathematical analysis introduces Laplace transforms, Bessel functions and

Laguerre functions without explanation, while a fairly common piece of equipment such as the speech spectrograph is explained for several pages. It is unlikely therefore that this book will be the one to put the new account of hearing prominently on the psychological map, while those psychologists who already work in the field may find that it is detailed primarily on points which they already know or which are outside their interests.

Nevertheless, this is an important work. Those interested in the area will be familiar with the idea that each cavity in the mouth contributes to the shape of the spectrum of the voice, and that the position of the peaks in that spectrum carry quite a lot of the intelligibility. The analysis of the vocal tract as a transmission line, however, implies that even two cavities will produce a whole series of peaks in the spectrum, rather than just two at the natural frequencies of the cavities. Furthermore, on this analysis it is not the case that the value of any one resonance is associated with any one cavity, so that the second resonance may be caused to rise either by a forward movement of the tongue, or in other cases by a backward movement of the tongue. Thus the production of the same change in a sound may result from quite different muscular movements. It is clear that some understanding of these relationships is essential for psychologists attempting to reveal the way in which articulation is controlled, and this in turn requires at least an outline understanding of the vocal tract considered as a transmission line. For this kind of reason, this book should be welcomed by all those interested in speech; but there is still room for a work more directly intended for a psychological audience.

D. E. BROADBENT.

Handbook of Mathematical Psychology. Edited by R. Duncan Luce, Robert R. Bush and Eugene Galanter. London and New York: Wiley. Vol. I. 1964. Pp. xiii + 491. 80s. Vol. II. 1964. Pp. 606. 90s. Vol. III. 1965. Pp. ix + 537. 90s.

As work on mathematical psychology multiplies there is an increasing need for formal teaching in this field at an undergraduate level; since there is no suitable textbook, the authors considered writing one, but ". . . there was too little preliminary organization of the somewhat scattered research literature to make the task feasible within a reasonable time," and therefore workers in the various fields of mathematical psychology were asked to contribute their own chapters. Little restraint was placed upon the contributors, and the result is far longer than had been anticipated, with some of the articles stretching to well over 100 pages. They are also difficult, with no special effort to meet the non-mathematical psychologist halfway, and the length and difficulty mean that only with drastic selection could it be used as an introductory textbook. The need for such a book still exists, but it would be most unreasonable to complain, since the present work is a godsend for the research psychologist who would like, from necessity or curiosity, to analyse his data and present his ideas in the language of modern mathematics. If psychology is largely mathematical in a few years time, this book will deserve much of the credit (or blame).

The first chapter is "Basic Measurement Theory" by Suppes and Zinnes, and their formal descriptions of empirical and numerical structures provide a logical basis for the chapters that follow. Most of these are classified according to the empirical areas with which they are concerned; they include, among others, "Detection and Recognition," by Luce; "Psychophysical Scaling," by Luce and Galanter; "Stimulus Sampling Theory," by Atkinson and Estes; three chapters on language by Chomsky and G. A. Miller; "Analysis of Some Auditory Characteristics," by Zwislocki; and "Preference, Utility, and Subjective Probability," by Luce and Suppes. In addition there is a chapter on computers in psychology by Newell and Simon, and an interesting but difficult chapter on parameter estimation by Bush. The two final chapters of volume III are straightforward mathematics, "Stochastic Processes," by Snell, and "Functional Equations," by Bellman; these are included because they deal with topics of obvious relevance, but for which there is no readily available introduction at the right level. There are some omissions, notably information theory and linear analysis of tracking behaviour; the editors explain that suitable introductions to these are obtainable elsewhere, but if this is the criterion it is difficult to see why several of the other chapters are included, particularly that on computers, on which there are plenty of general texts, and which is anyhow, as the authors point out (they begin: "The inclusion of a chapter about computers in a handbook of mathematical psychology is a historical accident.") scarcely mathematical. Information theory is not quite left out in the cold, since Miller and Chomsky

slip in a brief introduction to it in their chapter on "Finitary Models of Language Users."

The standard of argument and presentation is uniformly high, though there are differences among the authors in the mathematical sophistication assumed in the reader. In some chapters, such as Edwin Martin's "Concept Utilization," the argument is gone through step by step, and little is taken for granted, while in others, such as Sternberg's "Stochastic Learning Theory," the reader is clearly expected to have thoroughly absorbed the preliminary mathematical reading suggested by the editors in their preface. These differences in level of difficulty are not reflected in the ordering of chapters; thus Martin's is the last chapter on mathematical learning theory, but would make one of the best introductions, while Sternberg's is the first.

There is a more fundamental way in which the chapters of this work differ, and that is in the degree to which the mathematical imagination is chained to experimental data. At one extreme, for instance, there is "Detection and Recognition," by Luce, in which theories are developed in contact with relevant experiments, while at the other extreme there is "Identification Learning," by Bush, in which the main empirical restriction appears to be based upon what is intuitively acceptable. In the chapter by Luce the experiments would not have been performed, nor could their significance be understood, without the theories, and only the most hidebound empiricist will fail to be impressed; in Bush's chapter, on the other hand, the psychologist who relies for his excitement upon the fitting together of the empirical jig-saw puzzle, and requires nothing more from mathematics than a help towards doing this, will find little of interest (unless he can supply the empirical content himself). The fact that the editors make no distinction between these two extremes indicates that the spirit of the work is that of the mathematician and theorist rather than of the experimental psychologist. This is fair enough, as long as mathematical elegance and superficial plausibility do not lead to an acceptance of the truth of a theory in the absence of adequate empirical backing; the careers of Hull and Freud, and their followers, suggest that, even in the absence of elegance, this is a recurrent danger in psychology.

A lesser danger is that important problems, which do not lend themselves very easily to statement in mathematical terms, will be ignored in favour of problems that are, mathematically speaking, more tractable; one has a sense of this danger in the following quotation from Atkinson and Estes: "In order to provide rigorous and detailed tests of these concepts, it is frequently necessary to contrive special experimental routines in which the theoretical analyses generate tractable mathematical systems." It is worth noting that one of the most fully investigated problems of learning theory, that of the partial reinforcement effect, has not much occupied the attention of mathematical psychologists, and a recent stimulating book on the subject, by Lawrence and Festinger, was based on a theory that owed nothing to mathematics; an equivalent mathematical theory would have lacked the kind of flexibility that is often helpful to experimental progress.

Apart from these reservations, enthusiastic welcome seems the appropriate response for a work which should enable mathematics to penetrate into the many fields of psychological investigation where it can be of use. A. W. STILL.

Foundations of Psychological Research: Statistics, Methodology and Measurement. By Kenneth H. Kurtz. Boston, Mass: Allyn and Bacon. 1965. Pp. xiii + 402. $8.50.

The title is misleading, suggesting a grand book about logical foundations; instead "the aim is to provide some of the technical background necessary for an understanding of contemporary psychological theory and research." There are four parts, all at an elementary level. The first is a brief but carefully thought out account of the aims of experimental design; the second part, by far the longest, is on statistics, and in this the author attempts to convey insight, rather than to provide a cookbook; then there is a rather superficial section on scientific method, with a Skinnerian analysis of consciousness; and finally measurement and psychophysics. The content of each part is conventional, and the only originality lies in combining them to make a book; it would have been more interesting if there had been greater interaction between the four topics, instead of presenting them more or less independently. However, the presentation is clear, and it might prove a useful textbook for anyone trying to expand his elementary statistics course in the direction taken by Dr. Kurtz. A. W. STILL.

Statistical Methods: A Problem-Solving Approach. By Vivian Gourevitch. Boston: Allyn and Bacon. 1965. Pp. x + 310. $7.75.

For some time now we have been aware that the research behavioural scientist can hardly afford not to acquaint himself with statistical methods; the difficult question is how much mathematics is he supposed to learn in the process, or how much is he prepared to tolerate. Professor Gourevitch tries to present statistical techniques as being "firmly tied to sound methodology," and uses a wealth of illustrative examples and very little mathematical formalism. This extreme approach, to my mind, restricts the framework of this readable book and the reader is left without the tools for making simple extensions of the methods presented. However, with its account of IQ and Introversion tests and tests of personality theories, this book should be interesting to many.

E. A. C. Thomas.

Ergonomics: Man in His Working Environment. By K. F. H. Murrell. London: Chapman & Hall. 1965. Pp. xviii + 496. 63s.

Although not explicitly stated, this book is clearly intended for the man in the factory, be he manager, designer, draughtsman or work study engineer. While other people will undoubtedly want to read it and could profitably do so, the text primarily attempts to summarize . . . "such information as is at present available in a form in which it can be used to meet . . . human design requirements." However, an attempt has also been made . . . "to give the reasons for the conclusions drawn and recommendations made, so that those who may have to develop them further will know something of the scientific bases which underlie them."

The work is therefore divided into two main sections—a first part which occupies 124 pages and provides a general review of the structure, functions and limitations of the human body; and a second part which deals with practical ergonomics and contains detailed recommendations concerning design aspects, environmental factors and organizational considerations related to work and the work place. In general, this is a successful arrangement, although the allocation of topics between Parts 1 and 2 is somewhat arbitrary in many instances. This, together with a difference in depth of discussion of particular subjects has led to some anomalies; for example, in Part 1, the chapter on Metabolism and Heat Regulation (Ch. 3) occupies little more than six pages, while a discussion of the characteristics of the neurone and neuronal transmission itself takes up more than four pages of the chapter on the Nervous System (Ch. 5). Similarly, in Part 2, the two chapters on Design of Instrumental Displays (Ch. 9) and the Visual Environment (Ch. 14) together amount to approximately a third of the section which comprises 13 other chapters!

In a book on Ergonomics written by one person it is almost inevitable that some topics will be very much more effectively dealt with than others. The present work is no exception and clearly Murrell is at his best in the chapter on Design of Instrumental Displays (Ch. 9) which provides perhaps one of the most lucid and useful expositions of the subject. The many and various details appropriate to the choice and use of instruments are systematically treated and helpful consideration is given to the problems associated with instrument design. Unfortunately, the discussion of some other topics falls far below this standard as, for example, in the description of the Nervous System (Ch. 5). Here there are not only gross mis-statements of fact (e.g. that thinly myelinated fibres conduct nerve impulses faster than thickly myelinated ones) but also confusing inconsistencies and some surprising omissions. Furthermore, the loose usage of terms generally evident in this chapter is likely to make any self-respecting physiologist shudder!

On the whole however, the text is clearly written and well documented with over 480 references. There are ample illustrations and for the general industrial reader the book achieves its object and provides a useful source of information. For the research worker on the other hand, it does not go far enough and psychological issues, which are scattered throughout the text, tend to be very sketchily treated. The book is designed primarily for the British market and at 63s. the price seems reasonable enough.

K. A. Provins.

Psycholinguistics. Edited by Thomas A. Sebeok. Bloomington: Indiana University Press. 1965. Pp. xii + 307. $4.75.

This volume is a paperback reprint of three well-known works: Osgood and Sebeok's *Psycholinguistics—A Survey of Theory and Research Problems* (1954); a revised version of

Diebold's extensive review (*Language*, 1964) of Saporta's readings in psycholinguistics, and Miller's *The Psycholinguists* (*Encounter*, 1964). A good buy, then, in terms of value for money. But—the papers bridge a large time span (during which a major revolution has taken place) and are aimed at widely differing audiences. One may therefore question the wisdom of reprinting such a mixed bag under one cover without denying the value of the individual contributions.

The unifying theme of the Osgood and Sebeok section is the close relationship between taxonomic linguistics and behaviourist psychology. The notion of a K-limited stochastic source is used to explicate response-chaining of the units (on different levels!) isolated by taxonomic procedures. Diebold's paper is essentially an annotated bibliography of research between 1954 and 1964 and is thus a useful guide to an extensive literature. Miller lucidly outlines some of the psychological implications of transformational grammar. The discussion ranges from syntax (fairly well understood) to the obscurity of semantics and "belief" systems. Some of the implications of regarding complex organisms as "rule-governed" rather than "law-governed" are investigated. Students may find the variety of approaches discussed in the book rather confusing and could well be forgiven for imagining that the respective authors are talking about totally different subjects. A little editorial guidance would have been valuable if only to explain that in 1954 the goal of psycholinguistics was the prediction of verbal behaviour whereas a more meaningful aim would now appear to be the construction of a theory of linguistic performance.

JOHN C. MARSHALL.

Basic Psychology. By Howard H. Kendler. London: Methuen. 1965. Pp. xii + 750. 60s.

There are now so many introductory texts of psychology that recent authors feel compelled to justify their labours. Kendler's first aim is "to provide a realistic and clear picture of contemporary psychology." But there are many kinds of picture, and this one is clear in the way that a line diagram of a scene is clear; it concentrates on outlines, boldly presented, and leaves details, subtleties and difficulties to the imagination. For instance, the two chief rival theories of colour vision are discussed in one page, visual acuity in two, hearing in seven, the nervous system in 13. The book therefore whets the appetite, without even pretending to satisfy it. If by *realistic* is meant some account of every branch of psychology then the book is very nearly that. Its 17 chapters cover scientific method, the science of behaviour, origins of modern psychology, statistics, biological foundations of behaviour, sensation, conditioning, perception, motivation, learning and forgetting, verbal behaviour and problem solving, frustration and conflict, personality, behaviour pathology, social behaviour, psychological testing, and applied psychology. One can hardly quarrel with this list but the total omission of the related fields of information and communication theory, mathematical psychology, and psychophysics is a pity.

The author's second aim is "to meet the educational needs of introductory psychology courses," and here he scores rather well. There is sound advice on the nature of psychology in an introduction for students, the text is simply and clearly written, the frequent analogies are apt and incisive, and the whole is lavishly and clearly illustrated. There is an author and subject index, a useful glossary, suggestions for further reading after each chapter, and a more complete bibliography at the end. I only hope that the reader is not deterred by what is almost the opening statement of the preface, namely, "Psychology—a discipline that is hurtling towards scientific maturity—possesses an intrinsic structure that demands three-dimensional treatment." The book is a lot better than that, and one of its principal messages is that psychology has a long way to go, and that the psychologist cannot hurry.

A. COWEY.

Speech Disorders. By Lord Brain. London: Butterworths. Second Edition. 1965. Pp. vii + 201. 47s. 6d.

The first edition of this useful book was noticed in this *Journal*, 1963, **15**, 143. This edition contains a new chapter on current views on aphasia, in which attention is drawn to the more recent writings of Bay, Luria, Denny-Brown and (very briefly) the psycholinguists. The chapter on handedness and cerebral dominance has been revised to take account of recent studies using the intra-carotid sodium amytal technique, which has made possible for the first time a direct attack on the problem. The chapter on language disorders in children has also been appreciably expanded. Although it is impossible to include much discussion in so brief a compass, the conciseness achieved by the author

is particularly remarkable in a field in which lack of words has so often occasioned their profusion. This edition is a little more expensive than its predecessor but is 17 pages longer and its cover is gayer.

O. L. ZANGWILL.

Social Psychology: The Study of Human Interaction. By Theodore M. Newcomb, Ralph H. Turner and Philip E. Converse. New York: Holt, Rinehart and Winston. 1965. Pp. xii + 591. $8.50.

One great advantage of reviewing a textbook is that the reviewer is thereby made to read the text through, which is presumably an exercise normally performed only by students. The advantage of this is that "readability" is one excellent criterion for initial judgement, and this new text by Newcomb and his associates comes out of this assessment with high marks. This is a particularly deserved accolade in view of the increasing difficulty of presenting the material of social psychology in a coherent and readable form. In the halcyon days of 1950 when Newcomb published his first text, the field of social psychology was broad, yet still just about narrow enough to yield to a sweeping treatment. The material, too, was then in larger, more all-embracing, units which were susceptible to such treatment. The present text, however, focusses down on the psychological analysis of interpersonal situations without attempting to draw in more of the cultural and sociological complexities.

Thus we are now dealing with a completely new text which concentrates on one set of issues and makes no claim to be comprehensive. This is no bad thing, especially since the major thrust in recent social psychology has been within this area, and Newcomb has personally contributed a good deal. Inevitably there is a loss which the reader feels and especially in the implication of this type of study for the broader concerns of the world-at-large. This is partly a response to the fact that the ubiquitous laboratory animal—the American College student—inevitably figures prominently in a good many of its pages. This reviewer feels that one way to have remedied this situation within the authors' frame of reference, would have been to include some of the material arising out of the study of schizophrenic patients and their interaction patterns. Not all of this material would have been acceptable, but some of it has stimulated a good deal of work on interactional processes which is different from the more laboratory-centred research.

With this exception, the book presents a remarkably coherent view which manages to include a wide variety of material without strain, and synthesizes work in a way which has not previously been attempted on this scale. The days of *the* text in social psychology are now numbered, but within the field which it sets out to cover, this book can be recommended for all types of undergraduate course. As usual, the major barrier to its use is likely to be that of price.

DOUGLAS HOOPER.

An Experimental Approach to Projective Techniques. By Joseph Zubin, Leonard D. Eron and Florence Schumer. New York and London: Wiley. 1965. Pp. xix + 645. 105s.

To do this large and closely-packed book justice would require a long review. Briefly, its plan is as follows. An introductory chapter on "The Challenge of Projective Techniques" and an extended discussion of "Perception: an Approach to Personality?" lead into a systematic treatment of the two major techniques, Rorschach and TAT. With minor variations, the approach to each is the same: each is placed in its historical context, its current status is examined, and the principal methodological problems stated. Thereafter the technique is analysed *as an experiment*, i.e. the authors examine what happens when the Rorschach or TAT is administered, in terms of the elements essential to a scientific procedure: statement of an hypothesis, specification of apparatus and stimulus, etc. Next, they present their "psychometric approach," defining the variables involved and distinguishing between attributes of the stimulus and stimulus correlates, or determinants, of the response. The resulting "scoring systems" are very elaborate indeed; for Rorschach, 69 "scales" (many with sub-classifications) are proposed. For TAT, analysis is rather in terms of rating scales and check-lists. Finally, "derivatives" of each technique are briefly considered, a wider range, incidentally, from TAT than from Rorschach; Holtzman's technique is noted with approval as "superior to the traditional Rorschach for experimental purposes. . . ."

As indicated, the book is written from an avowedly experimental standpoint, and the documentation is exemplary. While it could go far towards resolving the question of the scientific status of projective techniques, it is unlikely to stimulate the interest of

the clinician, since it does not include the *Supplement* containing the tables, scales, charts and norms necessary to implement the use of the scoring systems. Lacking these, it is impossible to try the authors' ideas out in practice, and the almost complete absence in the book of illustrative material of any kind makes one chary of indulging in the additional capital outlay of $21.25 for photoprints (or $6.25 for microfilm), obtainable from the Library of Congress. BORIS SEMEONOFF.

On Binocular Rivalry. By W. J. M. Levelt. Soesterberg: Institute for Perception RVO–TNO. 1965. Pp. vi + 110.

This monograph presents experiments on both Binocular Brightness Mixture and Rivalry. It provides the best available summary of the literature on both topics up to 1963, being critical and quite comprehensive.

The author's own findings are as follows:

(1) The brightness of a binocular mixture is constant if a linear weighted sum of the monocular luminances is constant. In view of the large amount of physiological evidence suggesting that a non-linear transformation of stimulus energy is performed at an early stage of the visual system, this is an interesting and surprising result. One would like to see it verified for a larger range of brightnesses of the constant comparison stimulus than was used here.

(2) Binocular rivalry. Experiments are described on the effects on dominance and rate of rivalry of varying the size, blur, contrast or luminance of one of the rivalling stimuli. The results are in accord with earlier work, and the experiments are in general rather better. A new finding is that varying the blur, luminance or contrast of one stimulus alters the length of the periods for which that stimulus is suppressed in the rivalry, but not the periods for which it is dominant. This allows Levelt to give an attractively simple mathematical description of the relation of stimulus strength to dominance and rate of rivalry. However, again one would like to see the conclusion supported by data for a wider range of stimulus variables. Other experiments are described which demonstrate particular aspects of the well-known importance, in both phenomena, of contours or edges in the visual fields. There is a suggestive discussion of similarities between rivalry and the type of meta-contrast which depends on interaction between contours. Finally there is an ingenious analysis of the statistical distribution of the length of the phases of rivalry.

This work contributes more to showing when mixture and rivalry occur, and what their characteristics are, than to showing what features of the design of the visual system make them occur, and what advantages such a design has. But hardly anyone has yet attempted these more ambitious questions. PAUL WHITTLE.

Readings in Animal Behaviour. Edited by T. E. McGill. New York: Holt, Rinehart and Winston. 1965. Pp. x + 592. $12.00.

During the past two decades the study of animal behaviour has been rapidly emerging as a separate and coherent discipline within psychology and zoology. It differs from both comparative psychology and early ethology in the rounded view which it takes of animal behaviour. Thus the new "behaviourism" is concerned with six main questions—the immediate stimulus response relationships in any piece of behaviour, the mechanism (both theoretical and physiological) which underlies the behaviour, its genetic background, the development of the behaviour within the lifetime of the individual, the evolution of behaviour as revealed by comparative studies, and finally how the behaviour contributes to the survival of the species. It can be argued that this embraces the whole field of psychology, both human and animal, and that the main value of such a biological approach is to show the relationships and interdependencies between topics which have, in the past, been studied in sterile isolation.

When attempting to teach an undergraduate course based on this biological approach two main difficulties arise. First, there is no modern textbook which covers the area, and second, the original papers to which students are referred are scattered through journals in a field so wide that they are unlikely to be found in any one (or even three) departmental libraries. Students find it a very real problem to obtain the journals to which they are referred. In an attempt to meet this problem, T. E. McGill has edited a collection of original papers and articles published under the title *Readings in Animal Behaviour*. The papers all appeared between 1956 and 1961 and cover roughly those

topics listed in the first paragraph of this review. The selection is one of the best which the reviewer has yet met: indeed since it contains many of the references given jointly by Dr. Tinbergen and the reviewer in a first year course in comparative psychology it disarms criticism.

However, those criticisms which can be made are of two kinds. The first concerns the standard of the papers chosen. This varies from the chatty anecdotes of Breland and Breland (about failures to train animals with classical techniques) to the detailed analysis of brain chemistry by Krech and his colleagues. One is unable to follow up or assess the important issues raised by the first type of paper, and the second type requires a good background knowledge (e.g. in physiology or genetics). Thus while some papers are suitable for an elementary course others are not. A second criticism can be made with regard to the selection of papers, which seems inconsistent. It is strange that the section on genetics should omit any of the fine comparative studies of courtship in Drosophila species, subspecies and hybrids, while including four less intelligible and less elegant studies on mice. Similarly in the section on the neural and chemical control of behaviour there are three papers on hormones, two on brain stimulation, but none on electrophysiology (only a single paper in the whole collection) or brain lesions.

The editor has provided a brief statement at the start of each section indicating the main problems which arise in the field. The book would have gained coherence if these could have been expanded: at the moment they contribute little. The editor could indeed claim that he was not trying to produce a textbook, an assertion which only underlines our need for precisely that!

D. M. Vowles.

Human Factors Evaluation in System Development. By David Meister and Gerald F. Rabideau. London and New York: Wiley. 1965. Pp. xii + 307. 75s.

Some years ago, a major industrial group in this country realized, amazingly enough, that they had a human factors problem on their hands. They appointed a retired service officer to solve it for them; and he—honest man that he was—advised them to bring in a well-known firm of American consultants. It is saddening, and not a little alarming to reflect that better advice could hardly be given today. For the neglect of human factors work by British industry and, it is fair to add, by British academics, has been almost complete. It is unlikely that the publication of Meister and Rabideau's work will, by itself, do much to alter this; but at least, anyone interested can now consult this book to find out what human factors evaluation is all about.

The usual method of writing a book of this nature is to establish general principles, from which particular recommendations are deduced. The authors, however, have adopted a rather different approach. They examine in some detail an imaginary system, and their general ideas emerge as they proceed. The system they imagine is " 'Hermes' . . . an 18,000 lb., three-man, two and one half stage space vehicle designed to service and supply . . . satellites and space stations." This exacting requirement gives them plenty of scope, of which they take full advantage. They deal with the interacting stages of factors assessment, evaluation, and development, chronologically; and give proper weight to the constant pressure of the engineering and economic aspects of the total programme. It might be objected that they have selected an area where lavish facilities and expenditure are the rule; but this enables them to describe what ideally ought to be done: in any real case there would have to be compromises with what could be afforded.

The authors' purely psychological viewpoint seems a shade naïve, being largely limited to traditional S–R theory. Although decision taking is mentioned as a human function in an integrated system, the treatment it gets is surprisingly cursory: the reader will search the index in vain for the names of Tanner and Swets. Monitoring also receives very summary treatment; computer monitoring is not mentioned at all; there is no index entry for "vigilance." By contrast, the chapter on methods of data analysis contains a good deal of material which a reader could fairly readily look up elsewhere. Admittedly, the length of treatment each aspect of the subject requires is hard to determine, and there is no one "best" solution. Nevertheless, it seems to the reviewer that engineers have more need to be persuaded of the relevance of applied psychology than have psychologists of the uses of statistics. At least, it is surely so in this country.

With these reservations, however, the book is to be recommended; and it is to be hoped that it has much influence. The authors quote the cynical remark that ". . . human factors evaluations need not be performed because, even if performed, they are not allowed to influence system design." It would be a deserved success if they helped to falsify this aphorism.

M. Hammerton.

Visual Capabilities in the Space Environment. Edited by C. A. Baker. Oxford and London: Pergamon Press. 1965. Pp. vii + 203. 63s.

Now that the power and control engineering problems involved in getting several tons into orbital trajectories and back to Earth have been solved, the emphasis of Space Research has shifted to human life-support systems and how to make maximum use of human abilities in the space environment. As the Editor of this very useful collection of review articles puts it: "On a weight-volume basis, the human represents an economical addition of an adaptable computer whose primary input mechanism is the eye." These articles summarize what is known through laboratory experiments and the observations of American astronauts and Russian cosmonauts of visual performance in space. The 16 articles were sponsored by the Human Factors Society of America, the authors being actively engaged on human space flight. The answers are important for the space programmes, for tasks at present undertaken by special instruments will have to be taken over by humans for launches from alien worlds. Further, it is anticipated that the astronauts will be required to make essential visual judgements for docking and for moon landing.

It is unfortunate that many of the more systematic visual experiments have so far failed, for technical reasons, but the broad finding that human vision works well in zero gravity is of great interest. "Down is where the foot is," but there seems remarkably little confusion, at any rate for these highly trained men. John Glenn (in a quotation) has this to say of the reported remarkable acuity of space observers: ". . . I think you identify the smaller objects by their surroundings. For instance, you see the outline of a valley where there are farms and the pattern of the valley and its rivers and perhaps a town. You can see something that crosses a river and you assume that it is a bridge. As for being able to look down and see it and say that is a bridge, I think you are only assuming that it is a bridge more than really observing it." So, through the fallibility of perception, Earth becomes Space Man's fiction! R. L. GREGORY.

Color: A Guide to Basic Facts and Concepts. By R. W. Burnham, R. M. Hanes and C. James Bertleson. London and New York: Wiley. 1963. Pp. xviii + 564. 70s.

This book is unusual. It gives a great range of facts set out like Wittgenstein's *Tractatus*; written in so terse a manner that its statements take on an aphoristic quality as though it gives all there is to be said on the perception of colour.

Although primarily devoted to colour, it touches upon other aspects of vision, such as brightness and luminance units. These are given in extremely useful tables, allowing immediate comparison between the frightening number of alternative units and measuring scales. Essentially a reference book, it could be useful for students but it does not attempt to give any account of the various theories, with their merits and difficulties. For the teacher or research worker it is an invaluable handbook, giving almost all the known facts. The bibliography contains almost 300 references, but is far from complete, giving no mention of several important European investigators such as G. S. Brindley and W. A. H. Rushton. This is a sad lack, but nevertheless the book is a really useful addition to the literature on this confusing topic. R. L. GREGORY.

Visual Perception: The Nineteenth Century. By William Dember. London and New York: Wiley. Perspectives in Psychology Series. 1965. Pp. xii + 222. 23s. paper; 38s. cloth.

This is one of a series of readings, entitled Perspectives in Psychology, under the general editorship of William Kessen and George Mandler. The present volume offers a score of important, often exciting, extracts from the Founding Fathers of perception. It is a collection of Nineteenth Century gems, selected and cut with care by William Dember, who sets his chosen gems in useful historical comments.

Included are passages on the Eye, the Brain and the Senses by, respectively, Thomas Young, Charles Bell and Johannes Müller. We find here the original account of the Notorious Necker cube, Ernst Mach on contours, G. M. Stratton's re-inversion of the retinal image and Charles Wheatstone's wonderful paper on stereoscopic vision. When we come to the Master—Herman Ludwig von Helmholtz—we find him discoursing upon Unconscious Conclusions. It must have been difficult to select a single passage to represent Helmholtz, but the choice is wise for this issue is still controversial. We find, as usual with the Master, clear intimations of important later work—in this case the Ames

Room. "Looking at [a] room with one eye shut, we think we see it just as distinctly and definitely as with both eyes. And yet we should get exactly the same view in case every point were shifted arbitrarily to a different distance from the eye, provided all remained on the same lines of sight."

This was the great period of perceptual research, when the problems appeared fresh and were tackled by people who respected and understood other sciences. It was followed at the turn of the century by confused counterblasts, which left metaphysical smoke reeking in the air to cloud this first brilliant vision of how we see. R. L. GREGORY.

Visual Perception. By Herschel Leibowitz. The Critical Issues in Psychology Series; General Editor: Melvin H. Marx. London: Collier-Macmillan. 1965. Pp. xii + 177. 15s. paperback.

Readings in psychology are useful in providing students with original material uncontaminated by the textbook writers. They are especially useful in present conditions, where the student population stretches library facilities beyond reasonable limit. This book on visual perception is one of an unusual series, in which a collection of source papers is combined with a text in its own right.

The first half, entitled Inquiry and Argument, is a discussion of classical problems, including: perceptual adaptation, innate v. learning (rather inadequate), motivation, illusions, and the relation between perception and the nervous system, this being philosophical in tone. The author is an investigator of distinction, primarily on the development of the constancies in children, and the book is valuable for its treatment of this topic. Here Leibowitz writes with sufficient authority to set out the limits of what is known. Although there is but little reference made to underlying physiological mechanisms, there are interesting accounts of animal—including insect—behaviour studies and the whole is written with a commendable awareness of the biological importance of perception.

The Selected Readings given here amount to eight papers, including: the second of the Rock and Kaufman moon illusion studies and Held and Schlank on disarranged eye-hand co-ordination, which pairs well with a paper by von Holst. The selection is too small, and the book as a whole too short to do justice to the topic of Perception. But although an elementary book, it does illuminate some critical issues. R. L. GREGORY.

The Perception of Brightness and Darkness. By Leo M. Hurvich and Dorothea Jameson. Boston: Allyn and Bacon. 1966. Pp. ix + 141. $2.50.

This undergraduate text deals with the measurement of subjective brightness, thresholds, temporal factors, contrast and constancy. "Photochemistry and Neurophysiology" occupy a paragraph on the last page, although there is a considerable amount of quasi-physiological theorizing in the rest of the book. The imprecision of this theorizing, a tendency to avoid direct confrontation with whatever subject is under discussion, and defects of presentation, outweigh any virtues the book possesses. PAUL WHITTLE.

Modes of Thinking in Young Children: A Study of the Creativity-intelligence Distinction. By Michael A. Wallach and Nathan Kogan. New York: Holt, Rinehart and Winston. 1965. Pp. viii + 357. $8.00.

This book reports a cautiously ambitious research project concerned with psychometric assessment of features of cognitive function in 151 children aged 10 to 11 years. Assessment procedures are described at length for aspects of their behaviour in school and play settings, how they deal with certain categorizing, conceptualizing, and story-completion problems, their responsiveness to physiognomic properties of visual materials, their anxiety and defensiveness. Reliabilities of these measures are carefully considered and the overall concern is with statistical relations which emerge among these measures. The book's main value probably lies in these assessment procedures and their promise for future research but, for some readers, this value may be eclipsed by the fact that the report focuses on the topical and emotionally-charged distinction between intelligence and creativity. The authors start by reviewing the literature on this distinction as it applies to children and conclude that it has not been clearly established. They then propose several hypotheses about how "creativity" can, and cannot, be assessed and describe their own assessment procedures and the results of applying them to the 151 children. Their "creativity measures" show high reliability and inter-correlation, and so do their "intelligence

measures"; but the two sets of measures are substantially unrelated. At this point, they allot each child to one of four groups, that is, high intelligence-high creativity, high-low, low-high, and low-low. Subsequent measures are considered largely in terms of the average score for each of these four groups and the interpretation of these averages.

The findings and their discussion raise a great many points of detailed interest and hint at a coherent, but intricate, overall pattern. Above all, they raise many clear questions for elucidation by future research. Not least among these questions is the correctness of the authors' hypotheses about how "creativity" can be assessed. It cannot, they claim, be assessed by ratings from teachers, peers, or self; these claims remain open to experimental test. It can, they claim, be assessed by asking the child to engage in what might be called continuous controlled association or divergent production, e.g. name all the round things you can think of, or give meanings for this drawing: however, these tasks *must* be done in a context which is play-like and free from any hint of evaluation or time-pressure. Does the difference between a play-like and an exam-like setting radically alter a child's showing on these production tasks? Does the pattern of production really differ from one child to another in the way hypothesized on page 16? Can any child begin to score high on such a task if he is merely encouraged to persist with it? None of these questions is put to direct test, yet each is important for the overall thesis. These hypotheses may well be correct but it would be premature to conclude that they have been established by the many intriguing findings which are here reported.

I. M. L. HUNTER.

Bioelektrická Aktivita Kože (The bioelectrical activity of the skin). By A. Uherík. Bratislava Vydavatel'stvo Slovenskej akadémie vied. 1965. Pp. 197. Kčs. 23.50.

The problems of making a work known to a wider than national audience are fraught with difficulties. In this book a gallant attempt is made by giving the contents, the figure titles and summaries of the experimental findings in English and Russian for those who do not read Czech. The book is in two parts, firstly a review of the current position on the electrical activity of the skin and secondly an experimental section in which the author's own experiments are described. It is unfortunate that there is no English summary of the first part as the table of contents makes it clear that there are points of view which other followers of "the skin game" would be very interested to know about and which evidently refer to works only available in Czech. The experimental section which is fully summarized does not present particularly striking findings although some of those on the relations between stimulus intensity and features of G.S.R. response which are described have not been extensively examined in Western literature. In this subject where details of methodology tend to be very important in assessing results, summaries of findings while better than nothing, tend to be unsatisfactory and only serve to whet the appetite for more detail. Perhaps the most valuable aspect of this book is that it informs a western audience of particular aspects of the work that is going on and enables the relatively small numbers of workers in this field to add another member to their list of correspondents.

PETER H. VENABLES.

The Anatomy of Memory. Proceedings of the First Conference on Learning, Remembering and Forgetting. Edited by Daniel P. Kimble. Palo Alto, California: Science and Behavior Books, Inc. 1965. Pp. xi + 451. $4.35.

This "vinylback" volume records the transactions of a Conference held at Princeton in September, 1963, sponsored by the American Institute of Biological Sciences and organized by Dr. Karl Pribram. Much of it is concerned with what might be called the "microphysiology" of memory, i.e. the basic properties of neurones which permit functional modification. Sir John Eccles writes on physiological, Dr. Lawrence Kruger on morphological, and Dr. Holgar Hydén on molecular aspects. (This last is particularly fascinating.) There are also contributors on facilitation and impairment of storage processes (Dr. J. L. McGaugh), information transmission at synapses (Dr. A. M. Uttley) and "memory without record" (Dr. Heinz Von Foerster). The discussion is voluminous, often acute though sometimes tedious, and there has clearly been much secondary elaboration. Although the informality of the proceedings was doubtless genuine, in print the bonhomie appears not a little forced, the attitude of the contributors to one another being on occasion marked by an elephantine archness. It is pleasant to be able to record that three members and one former Visiting Foreign Member of the E.P.S. took part in this Symposium.

O. L. ZANGWILL.

THE QUARTERLY JOURNAL OF
EXPERIMENTAL PSYCHOLOGY

Vol. XVIII 1966 Part 3

LEARNING AND RETENTION OF WORD-PAIRS WITH VARYING DEGREES OF ASSOCIATION

BY

A. W. HEIM, K. P. WATTS, I. B. BOWER and K. E. HAWTON

From the Psychological Laboratory, University of Cambridge

The main aim of this experiment was to compare the role of degree of meaningfulness with that of frequency of repetition, in the learning and retention of word-pairs. The (student) subjects were divided into three groups. The first group learned 25 word-pairs whose members were frequently and highly meaningfully associated with each other. The second group learned 25 word-pairs which were associated far less frequently and meaningfully whilst those of the third group were as lacking in associative value as possible. The stimulus-words were ambiguous; they were identical for each of the three groups; the word-pairs were presented, in randomized order, repeatedly, during the learning stage until the subject achieved the criterion of 20 (or more) correct word-pairs. After an interval of 60–90 days, the subjects were retested for retention.

The results indicated that degree of meaningfulness plays a greater role than does frequency of repetition, both in learning and in recall; that the role of repetition tends to increase, however, as the meaningfulness of the material decreases; and that individual differences in method of learning increase with the difficulty of the material to be learned.

Introduction

Since experimental psychologists first concerned themselves with the study of learning and retention, frequency of repetition has been regarded as a major factor. Although the importance attached to repetition *per se* is perhaps somewhat less at the present time than it was in the days of Ebbinghaus (1885) and Thorndike (1911) few workers would disagree with McGeogh (1942) that frequency is a necessary condition of learning. In the present paper, an experiment will be described to test the hypothesis that degree of meaningful association plays a more important role than frequency of repetition in the learning and retention of paired word-associates. The experiment was designed also to investigate qualitatively the type of errors made during learning and in later recall, and to examine the relation between learning and retention ability, and between errors and forgetting.

Experimental Design

(A) Subjects

The subjects were 60 Cambridge University male science undergraduates, between 19 and 22 years of age. Thirty were biological scientists reading zoology, botany, agriculture or medicine, the majority reading the last. The other 30 were physical scientists reading physics, chemistry, mechanical sciences, electrical engineering, crystallography or mathematics. No psychologists were included as it was thought desirable to confine experimentation to "naïve" subjects. The subjects were divided into Groups X, Y and Z, as shown in Table I.

TABLE I

60 SCIENCE SUBJECTS

- 30 Biological
  - 10X
  - 10Y
  - 10Z
- 30 Physical
  - 10X
  - 10Y
  - 10Z

Thus each group contained half biological scientists and half physical scientists. Within the Z group, the biologists and physicists were also matched for university examination results.

(*B*) *Material*

The material to be learned consisted of one of three sets of 25 pairs of words. The stimulus words were ambiguous and were identical for the three groups, these 25 words being drawn from the Brook Reaction test (Brook and Heim, 1960; Hargreaves, Heim and Watts, 1963). The first set, X, was composed of highly associated pairs of words: all the response words in X assumed the meaning which had been most frequently understood in the Brook Reaction test, when this was given to a similar population. E.g.,

X	*stimulus word*		*response word*
	SCALES	⟶	WEIGHT
	BANK	⟶	MONEY

In almost every case, the X response word chosen by us was in fact the actual word most frequently given as a response in the Brook Reaction test. When this was not so, it was a case of our having had to choose between two or more response words of equal frequency, e.g. TRAIN → RAILWAY, or → ENGINE, or → TRAVEL, or → CARRIAGE (the last, itself an ambiguous word in this context).

The second set, Y, consisted of the same stimulus words as those used in X, but the response words were chosen such that the degree of association between stimulus and response was weaker than that in X. The Y response words had a recognisable association with a different, and more rarely attributed, meaning of the stimulus words. They had been offered in the Brook Reaction test but only infrequently. E.g.,

Y	*stimulus word*		*response word*
	SCALES	⟶	PRACTICE
	BANK	⟶	VIOLETS

The third set, Z, again consisted of the same stimulus words but the response words had no obvious association to the stimuli and had never been offered as a response in the Brook Reaction. E.g.,

Z	*stimulus word*		*response word*
	SCALES	⟶	BOOK
	BANK	⟶	EXAMPLE

The Z responses were selected from response words in X and Y, 12 from X and 13 from Y (see Appendix). By employing response words in Z which had appeared in either X or Y, the same range of word frequency was maintained in Z as in the other two lists. All of the 25 stimulus words were ambiguous; and, as illustrated above, the meaning of the X words, as implied by their responses, was in every case different from the meaning of the same words, as implied by the Y responses. In the Z list, the subject could of course infer whatever meaning he chose for the stimulus word, since the response word bore no immediate relation to the stimulus, however interpreted.

(*C*) *Learning*

In the learning stage of the experiment, one-third of the subjects learned the X word-pairs, one-third the Y and one-third the Z word-pairs.

In the first trial, or *the rehearsal trial* as it shall hence be called, the subject was instructed to learn the pairs of words as they were presented, typed on white cards. He was shown the stimulus word, typed on the left hand side of the card and was asked

to say this word aloud and to write it, immediately, in the left hand column of his three-columned answer sheet (see Table II). Then the response word, typed on the right hand side of the card, was exposed and this also the subject was asked to say aloud, and then to write down in the middle column of his answer sheet. The complete set of 25 word-pairs was presented in this way.

TABLE II

A TYPICAL TRIAL (Y)

First memory trial

article	clothing	apprentice
punch	hole	√
pass	friend	√
panel	judges	beating
spring	balance	√
strike	clock	√
analysis	character	report
character	feeling	letter
model	figure	example
scales	fish	practice
star	nebula	banner
swallow	insult	√
train	lion	√
figure	diet	√
record	file	√
angle	right	sight
score	music	√
race	mill	√
save	goal	√
party	drink	slogan
joint	rivet	√
service	car	√
print	book	frame
spirit	devil	gum
bank	violets	√

When the rehearsal trial was finished, the answer sheet was removed, the cards were well shuffled and presented again. In this second trial, or *first memory trial*, the stimulus word of each pair was again shown alone and the subject was requested to say it aloud and write it again on a fresh sheet of paper in the left hand column. Then he was asked to say and write down what he thought was the second member of the pair. When he had given his answer, he was shown the response word together with the stimulus word. If he had answered correctly, he placed a tick in the third column and was then shown the next stimulus word. If he had answered incorrectly, he read out the correct response word and wrote it in the third column, before proceeding to the next pair of words. In each case, the subject was encouraged to give a response word even if it was a guess, but he was never compelled to do so. He was also asked to cover up the previous lines of writing to ensure that he did not look back and get some hints as to the response words by elimination.

This procedure was repeated until the subject reached a criterion level of 20 (or more) pairs correctly learned out of the possible 25. An example of a typical first memory trial on Y is shown in Table II. The subjects were timed over each trial. No attempt was made to hide the stop-watch, but the tester emphasized that this was not to be regarded as a speed test. The subjects were not told that they were going to be retested after some weeks, but they were asked not to talk about the experiment to their friends as there was a possibility that the latter also would be tested.

(*D*) *Retention*

The subjects were tested for retention after a period of time ranging from 60 to 90 days. In the retention trial, the subject was shown each stimulus and was asked to give the response word, once again saying them aloud and writing them down. The subjects were not shown the response words this time. On this occasion it was stressed that they should make every effort to give a response to each stimulus word, even if this meant giving the wildest of guesses. Most subjects complied with this request although there was one notable exception who appeared to hate answering unless he felt sure his response was correct. At this stage the list of words was run through once only, and the trial was timed as before.

After this the subject was asked:

1(*a*) Whether he had discussed the experiment with anybody during the interval and, if so, when and with whom.

If the answer to 1 (*a*) was positive, he was then asked:

(*b*) Whether he had discovered that different lists of word-pairs were being used.

2. Whether he had suspected that he would be retested.
3. Whether he had consciously practised at all during the interval.
4. What means he had employed to learn the word-pairs in the first place.
5. How many correct responses he thought he had given in the retention trial.

Once again people were requested not to talk to their friends about the experiment. Half of the subjects were tested by Bower and half by Hawton. The same experimenter tested the same subjects during both learning and recall. No experimenter effect was found when the results of these two sub-groups were compared with one another.

Results

(*A*) *Comparison between Groups X, Y and Z*

Table III shows the mean number of trials taken to learn to the criterion level of 20, by each group. It shows also the mean number of correct responses given in the retention trial for each group. A Mann-Whitney U test (Siegel, 1956) was used to test the significance of the differences.

It may be seen that the subjects in Group X took significantly fewer trials to reach the criterion level than did the subjects in Group Y. Similarly, the subjects in Y took significantly fewer trials than those in Z. More word-pairs were retained after the 2–3 months interval in X than in Y, and more in Y than in Z, both these differences being significant. Thus, when degree of meaningful association (highest in X) is directly contrasted with frequency of repetition (lowest in X), it is clear that

the former produces the better retention (highest in X). On the other hand Z, which required the greatest number of trials to criterion (owing to the lack of meaningful association between Z's stimulus and response-words), yielded the lowest retention score.

TABLE III

COMPARISON OF X, Y AND Z IN RELATION TO (i) NUMBER OF TRIALS TO CRITERION AND (ii) MEAN RETENTION SCORE

Group	*Mean number of trials to criterion (incl. rehearsal trial)*	*S.D.*	*U*	*p*	*Mean retention score*	*S.D.*	*U*	*p*
X	2·25	0·43	32·5	<0·001	13·25	2·7	38·0	<0·001
Y	4·15	1·90	48·0	<0·001	6·70	1·14	112·0	<0·01
Z	9·40	4·24			4·10	3·70		

The mean retention score for Z was in fact pulled up because two subjects did surprisingly well, one (L) remembering 13 word-pairs and the other (M), 10. It is perhaps worth mentioning that M was the only one of the 60 subjects to answer positively on the question of conscious practice during the interval period.

The mean retention score on Y is perhaps slightly lower than would be expected. This may be explained by the fact that this list of word-pairs is not as easy as it might seem at first sight. Most of the Y subjects realized during learning that the word-pairs are associated but only weakly so. In view of the ambiguity of the stimulus words, this suggests varied possibilities in the retention trial. However, the situation in X and Z is relatively clear-cut in that the word-pairs in the former are meaningfully and frequently associated and those in the latter are as much lacking in association as we could ensure, given these experimental conditions. The X subjects stood a far higher chance if they were guessing at any of the response-words during any of the trials than did subjects in either of the other groups. As has been mentioned before, the associations in this group had been frequently given in the Brook Reaction test.

Figure 1 gives representative learning curves from the three groups, a typical subject being shown in each case. In addition there is the curve for one of the slower subjects in group Z. This particular case shows the characteristically erratic progress of such a subject. Initially this appears to be due to the lack of adoption of a consistent method of learning (see Section D below). The erratic progress in the later part of the curve may have been due to a fatigue effect. Subjects taking 16 trials over the learning phase were likely to have been in the test situation for almost 2 hr. Indeed some of the slower subjects spontaneously reported that they suffered from fatigue in the later stages.

(B) Group X: comparison of Brook Reaction test frequencies with (i) learning and (ii) retention

The Brook Reaction test frequency for each X word-pair, for a group of 150 science students, was tabulated. These response frequencies varied between 60 and 10 per cent. of the total number of responses from this group of 150. These

frequencies were compared in the present experiment, (i) with the number of times each pair was recalled correctly, on the first memory trial in X, (ii) with the number of times each pair was recalled correctly in the retention trial in X.

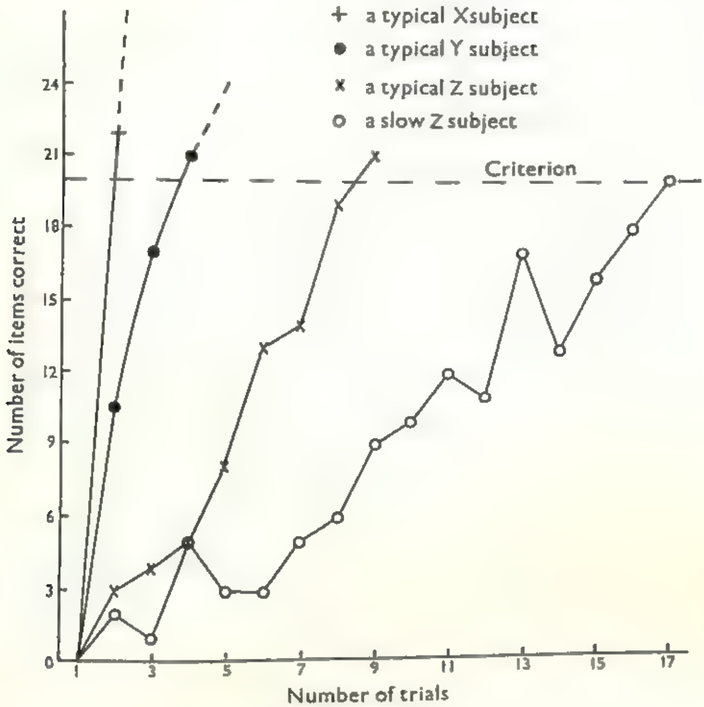


FIGURE 1
Some typical learning curves.

It may be seen that both correlations are significant: the higher the frequency with which a response word had been offered in the Brook Reaction test, the greater the ease with which that word-pair was learned (in one trial) and—achieving still higher significance—the better that word-pair was retained over the 2–3 month interval.

TABLE IV

COMPARISON OF BROOK REACTION TEST FREQUENCY WITH (i) FIRST MEMORY TRIAL AND (ii) RETENTION TRIAL

X LIST

	Correlation between the number of correct recalls and the Brook Reaction test frequency for each word-pair*	*d.f.*	*t*	*p*
Learning (first memory trial)	$\rho = 0{\cdot}352$	23	1·844	$<0{\cdot}05$
Retention	$\rho = 0{\cdot}426$	23	2·259	$<0{\cdot}025$

* Here, and throughout this paper, Spearman's ρ has been used. This was calculated by the exact method (Siegel, 1956).

This finding within group X merely underlines that described in Section A, where groups X, Y and Z were compared. The differences in degree of association are finer, however, and they are more quantitatively defined here.

(C) *Comparison of learning with retention*

The number of times that a pair of words was learned on the first memory trial was compared with the number of times that that word-pair was recalled at the retention trial, for groups X, Y and Z respectively (see Table V).

TABLE V

COMPARISON BETWEEN THE NUMBER OF TIMES A WORD-PAIR WAS RECALLED ON BOTH THE FIRST MEMORY TRIAL AND THE RETENTION TRIAL

Group	*Correlation*	*d.f.*	*t*	*p*
X	$\rho = 0{\cdot}483$	23	2·642	$<0{\cdot}01$
Y	$\rho = 0{\cdot}207$	23	1·014	N.S.
Z	$\rho = 0{\cdot}579$	23	3·402	$<0{\cdot}005$

Thus, in X and Z, those pairs of words which were most quickly learnt were significantly more likely to be retained. It should be mentioned that in X, there was rarely need for more than one memory trial, the majority of the word-pairs having been learned on the rehearsal trial. So when it came to ranking, the X score yielded a very narrow range. It is interesting that the correlation for Y is not significant: this is very likely due to the tendency of Y subjects to attribute an "incorrect" meaning to the stimulus-words.

If the number of repetitions of the material were the most important factor in retention, the subjects in each group who took more trials to reach criterion should have remembered more word-pairs in the retention trial, and any correlation between these two measures should be positive. In fact, the correlations for both Y and Z were low and insignificant. But whereas that for Y was negative ($\rho = -0{\cdot}17$) that for Z was positive ($\rho = 0{\cdot}13$). Thus the number of repetitions appears to have slightly more relevance as the material loses in meaningfulness. It should perhaps be recalled that subjects may have experienced some "cross currents" in the learning of Z, owing to the fact that the Z responses were drawn from X and Y response words and "scrambled."

The total time taken during the learning phase for each subject was compared with his retention score. Though none of the correlations was significant, there was an interesting trend. For X, the correlation was negative ($\rho = -0{\cdot}31$, $t = -1{\cdot}39$); for Y, $\rho = 0$; and for Z, the correlation was positive ($\rho = 0{\cdot}21$, $t = 0{\cdot}9$). Thus it would seem that there is some interaction between time taken to learn and degree of association: the faster learners tend to retain more in a task involving meaningful material whereas slower learners tend to retain more in a task involving less meaningful material.

(D) *Methods of learning*

After the retention trial the subject was asked what method he had used to associate the words in each pair. The methods employed were classified into the following categories:

FAMILIARITY—remembering because of the high frequency of previous experience of the association. For instance, most people habitually associate "bank" with "money" and therefore find this pair of words very easy to learn.

VISUAL—remembering the association by means of visual imagery. Thus one subject in Z remarked that he remembered SCALES → BOOK by visualizing a book sitting in one pan of a pair of scales. Another Z subject said that he remembered RECORD → MILL by means of a visual image of a record player in a mill house.

PARROT—rote meaningless learning, paying no attention to possible associations between words, if any.

FORCED ASSOCIATIONS—subjects in this category attempted to force unrelated words into some familiar or unusual association either by using connecting words or bizarre ideas.

FORM OF THE WORDS—remembering words by their shape or construction. Thus a subject might remember two words because they contained the same number of letters, e.g. STAR → MEAT (a Z pair) or because of common letters, e.g. TRAIN → INSULT (again Z). In choosing the word-pairs, however, the experimenters had tried to avoid the possibility of associations of this sort.

Table VI shows the number of subjects in each group using the various methods of learning. Where a subject said he had used two of these methods, he is represented by a half, under the two relevant headings.

TABLE VI

METHODS OF ASSOCIATION EMPLOYED

Group	*Familiarity*	*Visual*	*Parrot*	*Forced association*	*Form of the words*	*Unclassifiable*	*Total*
X	16·5	1·5	0	0	0	2	20
Y	11·5	5·5	1	0	0	2	20
Z	0	5·5	3·5	9	2	0	20

Some of the replies could not be satisfactorily classified. Subjects in X were bound to use their previous familiarity of the association of the words in each pair, in view of the way this list was constructed. It would not seem necessary to use "parrot-fashion" learning in Y since the words in this group are moderately associated. Nevertheless, two subjects in Y did report that they used this method for half of the pairs. (It is possible, of course, that they were not aware of the meaningful association between the words.) However, the "parrot-fashion" method was used far more in Z. Visual association was used equally often in Y and Z.

The majority of the subjects in Z appeared to try to force some meaning into their word-pairs in some way. Some of the slower subjects in this group seemed to start with no fixed plan for learning their pairs and half-heartedly adopted the parrot-fashion method. Eventually then seemed to decide that a definite policy would have to be adopted and then went on to make steady progress. These results, based on the subjects' introspective reports, did reveal that there are considerable individual differences in the method of learning employed. This finding is in general agreement with that of Carlson and Carr (1940).

(E) Errors

The errors made in X and Y during learning and retention were compared, to see whether there was any difference in the type of error made (see Table VII).

TABLE VII

COMPARISON OF X AND Y ERRORS IN (i) LEARNING, (ii) RETENTION, BASED ON THE MEANING OF THE STIMULUS WORD ASSUMED BY THE SUBJECT

		Number of errors in each group			
		X	Y	χ^2	p
Learning	Right Stimulus word meaning	33	65	>21	$\ll 0.001$
	Wrong Stimulus word meaning	54	350		
Retention	Right Stimulus word meaning	86	44	>20	$\ll 0.001$
	Wrong Stimulus word meaning	144	306		

The "right stimulus word meaning" signifies that the subject gave a response which, although mistaken, was in keeping with the meaning of the stimulus word chosen for the given group,

e.g. BANK → CHEQUE
instead of BANK → MONEY

The "wrong stimulus word meaning" signifies that the subject gave a response which was not in keeping with the meaning of the stimulus word chosen for the given group,

e.g. SCALES → FISH
or SCALES → JUSTICE
instead of SCALES → PRACTICE

The ambiguity of the stimulus words provided ample opportunity for mistaken associations. Table VII shows that the X subjects were far more likely than the Y subjects to associate to the "correct" stimulus meaning even when they were making an error. This supports what has already been said about the nature of Y. This comparison could not of course be made with Z, since in this group there was no possibility of inferring either a "correct" or an "incorrect" stimulus meaning.

The two groups were examined also to see whether the same sort of error was made on the retention trial as at the learning stage. It may be seen (Table VII) that the results in retention are in fact similar to those in learning.

(F) Comparison of first and second halves of each trial

It was noticed during the learning phase of the experiment that the subjects tended to give more correct responses during the first half of each trial than during the second half (see Table VIII).

If, for newly learnt material, memory obeys a fast decay function, we would expect subjects to score better on the first half of the subsequent trial than on the second. Most of the material in the first half of the trial will have been seen more recently than the material in the second half, despite the fact that the cards were

shuffled thoroughly after each trial. This difference would be accentuated in the first memory trial as this took much longer than did the previous ("rehearsal") trial. The results from Y are in agreement with this. Those in Z are less conclusive, although the same general tendency is found.

TABLE VIII

THE SIGNIFICANCE OF THE DIFFERENCES BETWEEN THE NUMBER OF CORRECT RESPONSES GIVEN IN THE FIRST AND SECOND HALVES OF VARIOUS TRIALS

Group	*Trials*	*d.f.*	*t*	*p*
X	no significant results	1		
Y	1st memory trial	19	2·92	<0·005
	2nd memory trial	18	2·82	<0·01
	All memory trials	48	3·83	<0·0005
Z	3rd memory trial	18	2·35	<0·025
	5th memory trial	15	2·53	<0·025
	All memory trials	115	2·40	<0·01
XYZ	All memory trials	189	4·32	<0·0005

An alternative explanation to these results can be given by an interference hypothesis, which makes difficulty dependent on serial position and accumulates during the trial.

(G) Experimental consistency

The experiment was examined in two ways for internal consistency.

(1) *Physical scientists and biological scientists.* Here the subjects were considered as two groups of 30, the one comprising biological scientists and the other, physical scientists. The differences in performance, both on the number of trials to criterion and on retention score, were compared for X, Y and Z respectively. In none of these was a significant difference found, although only Z subjects were matched for examination results.

(2) *Between the "talkers" and the "non-talkers."* Those who admitted talking about the experiment to other subjects and discovered that more than one group of words were being used, were compared with those who had not so talked. The mean retention score for the 10 "talkers" was 9·2, and that for the 50 "non-talkers" was 7·6. The probability of this difference occurring by chance was found to be 0·84 by a Mann-Whitney U test. Thus there is no evidence, either that talking affected the test results or that subjects who talked tended to be different relevantly from those who did not talk.

(H) Miscellanea

(1) *Time per trial.* The mean time taken per trial during learning and the number of trials taken to reach criterion were compared, to see if the slower learner, in terms of number of trials, took longer also on his individual trials. The results were insignificant although in Z a negative correlation was found ($\rho = -0{\cdot}267$). Thus in this group the subjects who took the least number of trials to reach criterion took longer per individual trial.

(2) *Examination results.* The tripos examination results of the subjects were compared with their number of trials to reach the criterion level and also with their retention scores. Small positive correlations were found in each case except in the learning of Y. It was noticed that two people in Z, who had obtained first classes in their tripos examination had the highest retention scores although they showed little superiority in the learning phase. These were the two subjects (L and M) who were mentioned in Results, A.

(3) *Estimation of retention score.* The estimated retention score given by subjects (after the retention test) was calculated as a percentage of the number actually correct. These percentages were 50, 65 and 88 for X, Y and Z, respectively. Thus, all three groups *under*estimate their retention score, but not to the same extent. The less meaningful the material, the more accurately the subject seemed to be aware of the number of correct responses he had given. There are a number of factors that could contribute to this result. Since, in Z, far fewer word-pairs were correctly recalled in the retention stage, this tended to reduce the number of possible scores which the subjects could reasonably say they had attained. In many cases they recalled only one or two, rendering it unlikely that they would make gross under-estimates of their score, as was possible in X and Y. Furthermore, subjects guessing at some of the associations in X would be far more likely to produce a correct response than if they guessed in either of the other two groups, since the response words in X had been frequently given in an association test. The subjects here might be aware of the number they had recalled, as opposed to guessed, and would therefore underestimate their retention score.

There were several cases in Z where subjects gave correct responses in the retention trial but afterwards said they had not remembered the association, although the response word appeared spontaneously when the stimulus word was seen in the retention trial. Such cases might have resulted from parrot-fashion learning, the stimulus word acting as a trigger to the response word without any meaningful association having been created between the two.

(4) *Correction of errors after time-interval.* The errors made during the last trial in the learning stage (i.e. that trial in which the criterion level was reached) were examined to ascertain whether there were any cases in which a response was given incorrectly but later given correctly in the retention trial. It was found that there were 18 instances of this in X, 10 in Y and only three in Z.

The relatively high X result is not surprising since, as remarked above, guessing in this group might well produce the correct response-word. Inspection of the curves in Figure 1, extrapolated beyond the criterion, shows that X will have learnt rather more than Y, and Y rather more than Z, during the last memory trial. Even if a subject gave an incorrect response on his last memory trial he still had a chance to learn the correct response as he was always shown it when going through the list.

Discussion

The aims of the experiment have been in general fulfilled. The results confirm the hypothesis that the degree of meaningful association between members of word-pairs plays a more important role than does frequency of repetition, both in learning the second member of the pair and in recall after some months: the more familiar the association, the fewer the number of memory trials—in some cases, one only—and yet the better the recall after 2 to 3 months. This was clearly shown as between groups X, Y and Z.

Within the Z group (the word-pairs which were more difficult since there was no obvious association between their members), however, there was a tendency for those subjects who required more learning trials to remember more word-pairs in the retention trial (see Results C). So it may be that in harder tasks, faster learning does not necessarily go with better retention. In any case, it would appear that the importance of the role of repetition is directly related to the meaningfulness of the material for the subject. The emphasis so often laid on repetition by experimental psychologists is perhaps explicable by the fact that (with notable exceptions, such as Bartlett (1932)) they have tended to confine themselves to material as lacking as possible in meaning.

With regard to their method of learning, not all the subjects found it possible to reply with certainty. Their introspection did, however, yield three points of interest. First is the increase in individual differences, as the task grew harder. As may be seen in Table VI, Z employed more methods than Y did and Y employed more methods than X did. Secondly, Y and Z made considerable, and equal, use of visual imagery, whereas X used this method scarcely at all. Thirdly, Z subjects evidently had often to contrive "forced associations," and sometimes to make use of the form of the words, whilst nobody in either of the other two groups used these devices.

It may be remembered that in this experiment, the Z response-words were selected from the X and Y responses, in order to ensure the same range of word frequency in Z as in the other two lists. Thus, this object was achieved very likely at the expense of producing some cross-currents in the minds of some of the Z subjects (see Results C). It is hoped, therefore, to follow up this experiment with one in which a Z′ group is given response words which again bear no obvious relation to their stimuli, but which—unlike the present Z list—also bear no relation to other stimulus words in the same list.

APPENDIX

Stimulus word	*Response word* X	Y	Z
analysis	chemistry	report	summer
angle	triangle	sight	drink
article	clothing	apprentice	lion
bank	money	violets	example
character	personality	letter	banner
figure	number	diet	rivet
joint	meat	rivet	mountain
model	girl	example	slogan
panel	wood	beating	cricket
party	drink	slogan	chemistry
pass	mountain	friend	ghost
print	book	frame	sky
punch	boxing	hole	bird
race	colour	mill	triangle
record	player	file	mill
save	rescue	goal	music
scales	weight	practice	book
score	cricket	music	violets
service	church	car	gum
spirit	ghost	gum	file
spring	summer	balance	apprentice
star	sky	banner	meat
strike	match	clock	clothing
swallow	bird	insult	rescue
train	travel	lion	insult

The two junior authors (I.B.B. and K.E.H.) did the testing and the analysis of the results as their research project for the Cambridge Part II Psychology Tripos course. The other two authors are indebted to the Medical Research Council for financial support.

References

BARTLETT, F. C. (1932). *Remembering.* London: Cambridge Univ. Press.

BROOK, D. F., and HEIM, A. W. (1960). A preliminary note on the Brook Reaction Test. *Brit. J. Psychol.*, **51**, 347–56.

CARLSON, H. B., and CARR, H. A. (1940). Rote and logical recognition memory. *J. exp. Psychol.*, **26**, 199–210.

DEESE, J. (1959). Influence of inter-item associative strength upon immediate free recall. *Psychol. Rep.*, **5**, 305–12.

EBBINGHAUS, H. (1885). *Uber das Gedachtnis: Untersuchungen zur experimentellen Psychologie.* Leipzig: Duncker & Humblot.

HARGREAVES, D. H., HEIM, A. W., and WATTS, K. P. (1963). An experiment on the effects of mental set in the Brook Reaction Test. *Brit. J. educ. Psychol.*, **33**, 236–9.

MCGEOCH, J. A. (1942). *The Psychology of Human Learning.* New York: Longmans, Green & Co.

SIEGEL, S. (1956). *Non-parametric Statistics for the Behavioral Sciences.* New York: McGraw Hill.

THORNDIKE, E. L. (1911). *Animal Intelligence.* New York: Macmillan.

Manuscript received 25th January, 1966.

VISUAL AND AUDITORY STORES IN SHORT-TERM MEMORY

BY

BENNET B. MURDOCK, JR.*

From Department of Psychology, University of Missouri

If retrieval in short-term memory can be either from a pre-perceptual sensory store or from a post-perceptual memory then recall should vary as a function of input into sensory store. To test this possibility two experiments with paired associates compared visual and auditory presentation under conditions as comparable as possible. In both experiments modality interacted with retention interval: more recency with auditory but, in Experiment 1, more primacy with visual. The interaction was taken as support for the hypothesis. An alternative hypothesis (that storage is post-perceptual but not a-historical) was discussed and weak negative evidence presented.

The general features of a model necessary to account for the experimental data on short-term memory gradually seem to be taking shape. At the very minimum it seems necessary to postulate one or more sensory stores, a central processing mechanism, and memory (Welford, 1960). Input through a single channel will pass through sensory stores and central processing mechanism into memory; a slow presentation rate will permit rehearsal (recirculation from memory through the central processing mechanism) (Broadbent, 1958), whereas a fast presentation rate may result in queueing and trace decay (Yntema, Wozencraft and Klem, 1964). One can also find evidence for rapid trace decay with overload (Sperling, 1960), or with multi-channel presentation (Mowbray, 1964). Memory may be a two-part system (Waugh and Norman, 1965), or not (Melton, 1963), but in any case the need for a distinction between "pre-perceptual" sensory stores and "post-perceptual" memory seems generally accepted (Melton, 1963; Mowbray, 1964).

The question to which this paper is addressed is whether retrieval is exclusively from memory. While an affirmative answer seems likely it is by no means necessary; it could be from sensory store and/or memory. Under presentation conditions normal for studies of short-term memory the persistence in sensory store is sufficient for normal processing, and the read-out is not necessarily destructive. So, at the time of the retention test, the trace may still be present in both sensory store and memory; two potential stores exist. While we know that sensory stores are vulnerable to overload and may decay quickly, we know relatively little about their mode of operation under more typical conditions.

It is proposed to test the locus of retrieval by comparing retention after auditory and visual presentation of verbal material. Paired associates were used, and they consisted of common English words. Presentation of a pair would provide the input to the sensory store, and the presentation rate was selected to allow processing of each pair immediately after presentation. The rationale behind this approach is as follows: It is assumed that sensory stores for auditory and visual differ, though how is not yet clear. Given that the subject is required to process each pair upon arrival it presumably enters memory. If retrieval is only from memory it should be a-historical; i.e. which sensory store it came from should be immaterial. However, if retrieval is from sensory store and/or memory, then input modality should not be immaterial. It would not be enough to show simple differences between auditory

* Now at the University of Toronto.

and visual because unequal processing demands might displace the entire curve; actually, one should predict an interaction of retention interval with input modality.

The reason for using auditory and visual presentation was that the same informational content could be presented in two different forms. Paired associates were convenient in that the input (i.e. each pair) was discrete and discriminable; thus input and processing could be alternated in a simple repetitive pattern. The presentation rate was at least sufficient for processing each pair as presented; any additional processing (i.e. rehearsal) could only work against modality differences. And, of course, the attempt throughout was to make conditions for auditory and visual presentations as comparable as possible. Two experiments were conducted, the first a "pure-lists" design (all pairs within a list either auditory or visual) with the second a "mixed-lists" design (both auditory and visual pairs within a list).

Method

Material. The stimulus material in both experiments consisted of common English words paired at random. All words were two syllables and not more than eight letters long; they were randomly selected from the Thorndike-Lorge (1944) G count of 20 and above except that homonyms, contractions, and archaic words were excluded. The two words in each pair were always presented successively regardless of whether mode of presentation was auditory or visual.

For visual presentation the pairs were photographed on 16 mm. film and projected by a Dunning Animatic single-frame projector on a screen approximately 10 ft. in front of the subject. For auditory presentation the pairs were recorded on an Ampex F-4460 Fine Line tape recorder and presented over an Ampex 2012 loudspeaker placed directly below the screen used for the visual display. For Experiment 1 (pure-lists design) the tape recording was made by a male speaker, and for Experiment 2 by a female speaker. While the pool of words was the same for the two experiments, the actual samples of pairs and lists were drawn independently in the two cases.

Procedure. In both experiments the presentation rate was 30 pairs/min. The duration of each of the words in the pair was approximately 0·5 sec. For auditory presentation the recordings were made by synchronizing with a metronome set at 120 beats per min., and for verification direct measurements of duration were later made by photographing oscilloscope traces of the wave form. For visual presentation the Dunning projector was simply stepped at the same rate (i.e. 120/min.). Thus, the pair of words required 1 sec. for display; this time was followed by a 1-sec. quiet period in which the subject verbalized the pair just presented. In effect, the experimenter presented A and B, then the subject repeated A and B for each of the A–B pairs in the list, and this repetition (vocalization) is taken as evidence of processing.

In general the procedure was similar to our previous experiments with paired associates (e.g. Murdock, 1963). After each list had been presented one of the pairs was tested; the A member of the critical pair was presented as the probe and the subject had to try to recall the corresponding B member. There was no delay between the last pair in the list and presentation of the probe. Naturally the critical pair was in no way distinguished prior to presentation of the probe. The subject was instructed to guess if not sure; omissions were not permitted. There was one testing session for each subject.

In Experiment 1 (pure-lists design) all lists were six pairs long, and there were 72 lists in all. The 72 lists were sub-divided into two sets of 36 lists each; tapes and film strips were made for each set. Half the subjects started with each set and with each mode of presentation; thus, with this 2 × 2 design, each list appeared equally often under auditory and visual presentation, and order of presentation was counterbalanced across subjects. The critical pair occurred once at each possible serial position in every block of six lists; the order within the blocks was randomized by selecting a different permutation for each of the 12 blocks.

In Experiment 2 (mixed-lists design) all lists were four pairs long, and there were 64 lists in all. All 16 permutations of auditory and visual (i.e. AAAA, AAAV, AAVA, . . ., VVVV) were tested four times each, once for each of the four possible serial positions of the critical pair. The order of the 16 × 4 = 64 conditions was randomized and was the same for all subjects. However, two sets of tapes and film strips were made which complemented each other; all pairs that were auditory on one set were visual on the

other set, and vice versa. Half the subjects were tested on each set. The probe was always presented in the appropriate modality. As in Experiment 1, no omissions were permitted.

To interleave the two modalities the tapes were recorded first; they were designed to run continuously and "blanks" were left for the visual pairs. Then, in making the film strips, the procedure was reversed and there the auditory pairs were omitted. During the experimentation the Dunning projector was stepped by a pair of re-cycling Hunter timers; the timers were simply started and stopped by the experimenter as necessary. The (120/min.) metronome beat used for timing the auditory presentation was recorded on the tape along with the word pairs; for synchronizing of auditory and visual pairs during experimentation the experimenter simply started and stopped the timers in time with the metronome. In general the synchronization seemed adequate and little trouble was encountered during experimentation.

Subjects. All subjects were from the Missouri student pool, undergraduates of both sexes who were fulfilling a requirement of the introductory course. They were tested individually, 20 subjects being used in Experiment 1 and 36 in Experiment 2.

Results

Difficulties in processing the pairs were infrequent. In Experiment 1 no record was kept. In Experiment 2 almost half (specifically, 17 out of 36) the subjects verbalized all 256 pairs correctly and, over all subjects, the error rate was about 2·5 per 1,000 pairs. The errors that did occur were equally divided between auditory and visual presentations.

Although the processing errors were infrequent they were suggestive. All but one were either A–$\bar{B}$ or B–$\bar{A}$, where order of verbalization is as indicated and the bar denotes an incorrect response. Output interference (Tulving and Arbuckle, 1963) is suggested by A–$\bar{B}$; while speaking A, B is forgotten. Input interference is suggested by B–$\bar{A}$; during presentation of B, A is forgotten and the subject reports first what he remembers. While cases of the latter outnumbered the former by more than 2:1, the more surprising fact seems to be that both input and output interference may be manifest in the presentation and processing of a single paired associate.

The main results of Experiment 1 (pure-lists design) are shown in Figure 1, where recall probability is plotted as a function of serial position (of the probe, that is). As can be seen, the auditory curve shows more recency but less primacy than the visual curve. A three-way (modality $\times$ serial position $\times$ subjects) analysis of variance of number of correct recalls showed that the relevant interaction (modality $\times$ serial position) was highly significant; $F\ (5,95) = 4{\cdot}90$, $p < 0{\cdot}001$. While the effect of serial position was also highly significant, the modality comparison was not; $F\ (1,19) = 2{\cdot}34$, $p > 0{\cdot}05$. Thus, there was no overall difference between auditory and visual but a significant interaction with retention interval was obtained.

Not only were total number correct about the same for auditory and visual but also the absolute numbers of extra-list and intra-list intrusions were essentially identical for both modalities. However, the distribution of intra-list intrusions across serial position was not; instead, it showed exactly the same trend as correct responses. That is, the relative frequency of intra-list intrusions from position 5 was clearly higher for auditory than for visual, but for intrusions from positions 1–3 was higher for visual than for auditory. (The cumulative distributions across serial positions differed at the 0·05 level by a Kolmogorov–Smirnov two-sample test.) Thus, two independent measures showed the same interaction. For correct recalls auditory presentation was better than visual at the end of the list but visual was better than auditory for the beginning of the list. Given that an intra-list intrusion

occurred, it was more likely to come from the end of the list with auditory presentation but more likely to come from the beginning of the list with visual presentation.

The main results of Experiment 2 (mixed-lists design) are shown in Table I, which gives recall probability as a function of serial position for auditory and visual pairs. As can be seen auditory was generally superior to visual, but the differences were greatest at the two middle serial positions. The same type of three-way analysis

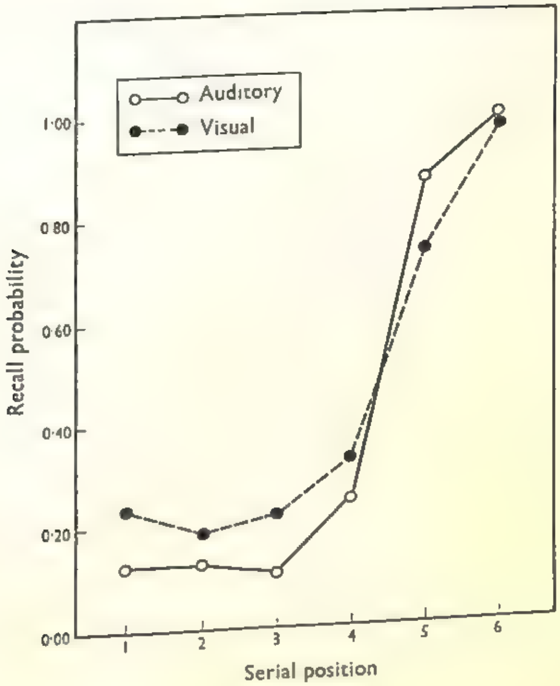


FIGURE 1

of variance as used in Experiment 1 showed that the effects of modality, $F(1,35) = 38 \cdot 3$, and the interaction of modality × serial position, $F(3,105) = 10 \cdot 2$, were both significant at the 0·001 level. Thus, the relevant interaction was again significant.

In Experiment 1 about two-thirds of the errors were intra-list intrusions; in Experiment 2 it was about seven-eighths. In so far as analysis was possible, it can be said that in the second experiment the intrusions did not show a greater recency effect for auditory than for visual. This analysis could reasonably be based on only

TABLE I

Recall Probability as a Function of Serial Position for Auditory and Visual Pairs

Experiment 2

Position	*Auditory*	*Visual*
1	0·392	0·316
2	0·601	0·399
3	0·896	0·694
4	0·986	0·976

16 lists: (*a*) those with the test pair in position 1 and ipsimodal distractors in positions 2 and 3 (i.e. the pairs in positions 2 and 3 either both visual or both auditory); or (*b*) those with the test pair in position 2 and ipsimodal distractors in positions 1 and 3. Then, intrusions from the two distractor locations could be dichotomized as "early" or "late." If anything, the relative incidence of "late" intrusions was greater when the two distractors were visual than when they were auditory. We have no ready explanation for why there should be a preponderance of "late" intrusions for auditory in Experiment 1 but not in Experiment 2.

A further analysis showed that the intrusions in Experiment 2 clearly tended to come from the same modality as the test pair. This analysis was based on another subset of pairs; specifically, those with test pairs at positions 1 and 2 and contramodal distractors (at positions 2 and 3 in the first case, at positions 1 and 3 in the second). For both visual and auditory test pairs, the modality of the intrusion was clearly appropriate; a 2×2 table (modality of test pair by modality of intrusion) showed non-independence with $\chi^2 = 46 \cdot 4$.

Discussion

The results of these two experiments clearly indicate that there are differences resulting from auditory and visual presentation of paired associates, even when each pair is processed immediately after presentation. Moreover, modality interacts with retention interval in such a way that auditory presentation clearly gives better recency than visual presentation; for primacy, either visual is better (Experiment 1) or they do not differ (Experiment 2). The same general sort of interaction has been suggested by McGhie, Chapman, and Lawson (1965) though they did not use paired-associate material.

The implication of these findings would seem to be that retrieval can be either from sensory stores or from memory. Certainly storage cannot be completely a-historical or modality would not have had such a large effect. It was, after all, exactly the same material informationally; it was merely presented in two different ways. This same conclusion was also suggested for the results of a study of forward and backward associations (Murdock, 1966). If retrieval is at least partly from sensory stores then associative symmetry would be expected, since the input of the two members of the A–B pair was always simultaneous. It is only with processing that order relations are introduced.

An alternative explanation for the present results would be that retrieval is exclusively from memory but visual and auditory inputs are stored with different tags. Thus, recall would be post-perceptual but not a-historical. As a weak test of this alternative, a supplementary experiment was conducted with voice quality rather than modality as the main experimental variable. If there is tagging in storage, then it should be demonstrable with other dimensions. The experiment was an exact duplication of Experiment 2 in all respects except that male and female speakers were substituted for visual and auditory presentations. The results were completely negative; not only were there no differences in accuracy but also the intrusions were completely symmetrical. Clearly, then, the alternative of differential storage is not supported by this supplementary experiment.

If one looks at the penultimate pair in the list it is abundantly clear that its recall is better when presentation is auditory rather than visual. This difference is not due to differences in degree of original learning; in Experiment 1 the early pairs in the list were *better* remembered for visual than for auditory, hence they must have been "learned." In Experiment 2 an analysis of only those subjects showing

perfect recall of the last pair in the list showed that they were still, by a sign test, better when the penultimate pair presented was auditory than when visual (22/3, $p < 0{\cdot}001$). It could be that visual pairs decayed more rapidly than auditory, or it could be that there were differences in interference generated by the last pair in the list. These two alternatives can easily be separated in Experiment 2, and it turns out that the former is more nearly correct than the latter. That is, retention of the pair at position 3 was completely independent of the modality of the pair at position 4; $\chi^2 < 1{\cdot}00$. However, whether the modality effect is due to differences in storage or differences in retrieval is not quite so easy to determine.

This investigation was supported by PHS Research Grant MH 10,882 from the National Institute of Mental Health, Public Health Service. The writer would like to thank Warren Carlson and Anne Stricklin for testing subjects and analysing data.

References

Broadbent, D. E. (1958). *Perception and Communication.* Oxford: Pergamon.

McGhie, A., Chapman, J., and Lawson, J. S. (1965). Changes in immediate memory with age. *Brit. J. Psychol.*, **56**, 69–75.

Melton, A. W. (1963). Implications of short-term memory for a general theory of memory. *J. verb. Learn. verb. Behav.*, **2**, 1–21.

Mowbray, G. H. (1964). Perception and retention of verbal information presented during auditory shadowing. *J. acous. Soc. Amer.*, **36**, 1459–64.

Murdock, B. B., Jr. (1963). Short-term memory and paired-associate learning. *J. verb. Learn. verb. Behav.*, **2**, 320–8.

Murdock, B. B., Jr. (1966). Forward and backward associations in paired associates. *J. exp. Psychol.*, **71**, 732–7.

Sperling, G. (1960). The information available in brief visual presentations. *Psychol. Monogr.*, **74** (11, Whole No. 498).

Thorndike, E. L. and Lorge, I. (1944). *The Teacher's Word Book of* 30,000 *Words.* New York: Teachers College, Columbia Univ.

Tulving, E., and Arbuckle, T. Y. (1963). Sources of intratrial interference in immediate recall of paired associates. *J. verb. Learn. verb. Behav.*, **1**, 321–34.

Waugh, N. C., and Norman, D. A. (1965). Primary memory. *Psychol. Rev.*, **72**, 89–104.

Welford, A. T. (1960). The measurement of sensory-motor performance: Survey and reappraisal of twelve years' progress. *Ergonomics*, **3**, 189–230.

Yntema, D. B., Wozencraft, F. T., and Klem, F. T. (1964). Immediate serial recall of digits presented at very high rates. *Paper presented at Psychonomic Soc.*, Niagara Falls, Ontario.

Manuscript received 17th December, 1965.

REMEMBERING A LIST OF TWO-DIGIT NUMBERS

BY

H. C. A. DALE and A. D. BADDELEY

From the Medical Research Council, Applied Psychology Research Unit, Cambridge

The way subjects remember a list of two-digit numbers has been examined in some detail. It is found that intrusions in free recall are not random. They resemble omissions in having the same first digit but not in other ways. This non-randomness of recall errors has been used to construct recognition tests of varying difficulty. Numbers which occurred commonly as intrusions were difficult to distinguish from the correct items when used as distractors in recognition tests. The experiments suggest that the previously observed relationship between recognition efficiency and number of alternatives (Davis, Sutherland and Judd, 1961) can be attributed to the increased probability that such intrusions will be included when the total number of distractors is increased.

Introduction

Davis, Sutherland and Judd (1961), using a list of two-digit numbers as stimuli have shown that if memory is tested by requiring the subjects to pick out the stimuli from an array which includes other two-digit numbers as distractors, performance deteriorates as the number of distractors is increased. When all possible distractors are included the number of stimuli correctly selected is roughly the same as the number which can be produced in free recall. This finding, which we have been able to replicate, can be interpreted, as Davis *et al.* have done, on the grounds that the amount of information retained by the subject is constant and that relatively high performance is possible on the basis of this constant amount of retained information where the number of distractors is small. This approach is valuable since it has long been known that the difficulty of recognition tests is variable and information theory promises to provide a rationale for one way in which it may be systematically adjusted. Two limitations of this use of information theory have to be recognized, however: (i) the simplifying assumption that recall errors are random is unlikely to be true as Davis *et al.* themselves acknowledged, and (ii) any such blanket measure provides no indication of the way in which the data have been stored and retrieved.

Brown (1959) has likened the decay of a memory trace to the gradual erasure of a chalk mark which makes the mark less distinct and hence increasingly difficult to read. Although uncertainty will in general increase, the confusions which will arise will not be random and will depend upon the form in which the data have been coded.

In the experiments to be reported it is first demonstrated that errors in free recall are non-random. This fact is then used to construct recognition tests in which the difficulty depends upon the nature of the distractors, their number being held constant. A preliminary report of a part of this work was published previously (Dale and Baddeley, 1962). The remaining experiments use intrusion errors to investigate the form in which the memory trace is coded. Two techniques are used: (i) the detailed analysis of errors in free recall; and (ii) varying the relation between correct items and distractors in recognition tests on the grounds that rejection of the distractors will be difficult if, when coded, they resemble the correct items.

Part I. Errors in Free Recall

Method

Stimuli. The following 15 randomly selected two-digit numbers were used: 11, 13, 26, 28, 32, 33, 36, 42, 51, 58, 59, 68, 83, 95, 97.

Procedure. Each stimulus was printed in black ink from 3·5 in. stencils on a 10 × 24 in. white card. The cards were exposed for 5 sec. with an interval of 1 sec. between each. Two different randomly selected orders of presentation were employed. Four groups of approximately 25 subjects were tested, two with each order. Immediate free recall was demanded. Subjects were told in advance that 15 numbers would be exposed and that the order of presentation did not have to be remembered. They were reminded of both facts while recalling.

Subjects. One hundred young enlisted men.

Results

The mean number of correct responses per subject was 9·66, while the mean total number of responses made by each subject was 13·75.

An analysis of the intrusions is given in Table I from which it is clear that some numbers, e.g. 98 and 57, occurred frequently whereas others were never given.

TABLE I

THE FREQUENCY OF EACH POSSIBLE INTRUSION

		Units									
		0	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*
Tens	*10*	0	×	1	×	7	8	3	2	6	0
	20	0	6	11	11	7	6	×	3	×	1
	30	0	7	×	×	2	7	×	6	18	4
	40	0	2	×	12	3	5	11	5	18	2
	50	0	×	12	8	4	5	14	24	×	×
	60	0	0	12	7	10	8	0	8	×	3
	70	0	3	3	2	2	1	2	0	4	2
	80	0	2	6	×	3	18	17	14	3	2
	90	0	2	4	6	1	×	6	×	20	0

× indicates correct response.

The χ^2 test provides an index of the non-randomness of the distribution of errors based on the assumption that each subject's intrusions are independent. This shows an extremely high level of significance ($\chi^2 = 200$ for 74 *d.f.*, $p < 0{\cdot}001$). An objection to this calculation is that different subjects contribute unequally. A derivation of chi-squared by Stone (1960) avoids this difficulty. To use Stone's method the first three intrusions of the 72 subjects who made at least three intrusions were considered. The distribution of these was markedly non-random ($p < 0{\cdot}001$). One implication of this result is that to some extent the subjects must have remembered this material in the same way. If each had employed a highly idiosyncratic storage and retrieval process or strategy the distribution of intrusions would have been random.

PART II. SELECTED RECOGNITION DISTRACTORS

It is possible that the intrusions in free recall bear no specific relation to the stimuli. On the other hand they could be phenomenally similar to omissions and indicate partial forgetting. If this latter is the case, then common intrusions should be relatively difficult to discriminate from the stimuli. They should therefore be more effective as distractors in a recognition test than numbers which never or rarely occur as intrusions. This hypothesis is tested in the experiment to be reported.

Method

Procedure. The 15 two-digit numbers used in Part I were presented under the same circumstances and in the same two orders to similar subjects. Immediate memory was tested by recognition tests in which the stimuli together with 15 distractors were printed in three columns. Printed instructions reminded the subjects that 15 numbers had been presented. In Experiment A only they also included the statement "If you are not sure, guess."

The distractors were systematically selected from the free recall intrusion data (Table I). In Experiment A two extreme sets were obtained by selecting on the one hand the 15 numbers which occurred most frequently as intrusions and on the other hand the 15 which least frequently occurred. The rare intrusions consisted mostly of multiples of 10 and 11 and were readily identifiable as a group. For this reason Experiment B was conducted in which multiples of 10 and 11 were excluded. The common intrusion distractors were the same in both experiments, with one minor exception (a substitute had to be provided for a multiple of 11).

The subjects were tested in groups. Recognition test sheets of the two different kinds were distributed so that within each group alternate subjects had different forms of the test. In Experiment A two groups of 18 subjects were run, one with each order of presentation. In Experiment B, four groups of 10 were run, two with each order of presentation. Test sheets were left face downwards on each subject's desk during presentation. They were turned over when a signal to respond was given at the end of the presentation.

Results

Mean performance levels under each condition and a summary of the statistical tests carried out are given in Table II.

TABLE II

MEAN NUMBER OF CORRECT ITEMS SELECTED BY EACH SUBJECT ON RECOGNITION TESTS WITH RARE OR COMMON ERRORS AS DISTRACTORS TOGETHER WITH THE MEAN NUMBER OF RESPONSES MADE BY EACH SUBJECT

	Experiment A (Extreme sets)			*Experiment B (Multiples of 10 and 11 excluded)*		
	n	*Mean score*	*Mean No. of responses*	*n*	*Mean score*	*Mean No. of responses*
Common errors	18	10·44*	14·94	20	10·75†	14·85
Rare errors . .	18	13·06*	14·39	20	12·05†	15·00

* $t = 4{\cdot}8$, $p < 0{\cdot}001$ † $t = 2{\cdot}6$, $p < 0{\cdot}02$

A marked difference between the two recognition tests was found in Experiment A; common intrusions constituted relatively difficult distractors. In Experiment B the difference was smaller but was still significant.

The implication of the differential effect of the sets of distractors is twofold. Since the technique has involved the comparison of data collected from different groups of subjects it confirms that these subjects must all remember this material in a similar way. It also indicates that there is some communality between the processes of recognition and recall.

PART III. REPLICATION OF THE RECOGNITION EXPERIMENT

To check the generality of the finding that recognition tests of varying difficulty can be constructed on the basis of intrusion errors in free recall the experiment has

been repeated using new stimuli, subjects from a different source and a slightly different experimental technique.

Method

Procedure. A fresh set of 15 numbers was used. No numbers used previously were included, nor were multiples of 10 and 11. Immediate free recall was obtained from a large number of subjects and intrusions were analysed so that high and low frequency intrusions could be selected as recognition distractors. No multiples of 10 or 11 were used as distractors. The procedure was the same as in Parts I and II above with the following exceptions:

(i) Subjects were tested in groups of 10 to 20.

(ii) A new order of presentation was used for each group. This meant that the orders used in the recognition tests differed from those used in the free recall tests.

(iii) The orders of presentation were random with the constraint that ascending or descending triplets were excluded.

Subjects

(i) Free recall, 10 young enlisted men plus 83 members of the APU subject panel of paid volunteers (these were predominantly housewives between the ages of 30 and 50).

(ii) Recognition, 69 members of the APU panel.

Results

With common intrusions as distractors the mean number of correct items selected per subject was 9·71 (S.D. = 1·46). With rare intrusions it was 10·76 (S.D. = 1·45). This difference is statistically significant ($t = 3{\cdot}24$, 81 *d.f.*, $p < 0{\cdot}01$). Thus the previous findings are confirmed.

The relatively small size of the effect can in all probability be attributed to the variation in the order of presentation. It might reasonably be assumed that intrusions are alternatives to items in the list with which they are subjectively equivalent at the time of recall. Common intrusions constitute difficult distractors in recognition tests because this same subjective equivalence is operative. In general, numbers at the beginning and end of the lists are remembered accurately. For a given order of presentation the subjects' memory for the middle items will be imperfect and the intrusions will be alternative to these. If the same order is used in recognition, then the effect of the difficult distractors is maximized since these are subjectively equivalent to the middle items. By varying it, their effect is minimized, hence the relatively small difference obtained in this replication.

Part IV. A More Detailed Examination of Intrusions in Free Recall

It was hypothesized above that the intrusions were alternatives to omissions, but no supporting evidence was presented. In the present section three hypotheses concerning the nature of intrusions are examined: (i) they are relatively highly available responses within the category of two-digit numbers; (ii) they are similar to omissions and might therefore be alternatives to them; (iii) they are similar to items correctly recalled and might be generated on the basis of these. These three hypotheses are examined in turn.

(i) The response availability hypothesis

If common intrusions are relatively available responses whereas rare intrusions are relatively unavailable, lists composed of the former should be easier to learn than lists composed of the latter. This has been tested directly.

Method

The two sets of numbers used as distractors in recognition experiment B above were used as two separate lists to be learned. They were presented as in Section I and immediate free recall was demanded.

Subjects. Seventy-four fresh young enlisted men tested in groups of 10 to 20. Thirty-seven were given each list.

Results

The mean number correct per subject was: common errors 9·35; rare errors 8·86. The small difference is not statistically significant ($t = 1{\cdot}06$ with 72 *d.f.*, $p > 0{\cdot}25$). It can therefore be concluded that response availability is not an important feature of the intrusions.

(*ii*) *The relation between intrusions and omissions*

Procedure 1. Data from one group in the original free recall test (cf. Part I) were analysed in detail. For each subject (N = 24, because one subject made no intrusions) the total number of common digits in the correct position was calculated (thus if he omitted 83 and gave 86 one point would be added, likewise if he gave 93, but none would be added for 98 or 31). The number to be expected by chance was calculated by considering all possible intrusions, counting up the common digits and deriving the mean number per intrusion.

Results

The mean number of common digits per intrusion was 1·56, whereas the mean number expected by chance was 1·16. Nineteen of the subjects had more common digits than would be expected by chance. Whereas only four had fewer. These proportions are significantly unequal ($p < 0{\cdot}001$), so it can be inferred that there was some overall resemblance between omissions and intrusions.

Procedure 2. This was used to see whether the first and second digits contributed equally to this similarity. Data were taken from two other groups in the original free recall test which yielded 49 usable records, and also from 62 of the subjects used in the replication experiment. The form of analysis was generally the same as that used above. Each digit was considered separately, so it was possible to count up for each subject (*a*) the number of intrusions which had the same first digit as each omission; (*b*) the number which had the same second digit; (*c*) the number of intrusions for which the second digit was the same as the first digit of each omission, and finally (*d*) the number of intrusions for which the first digit was the same as the second digit of each omission. As a baseline, the number of coincidences expected by chance was calculated by considering the particular omissions made by each subject. The overall mean number of coincidences could then be compared with the mean number expected by chance, and also the number of subjects who individually showed a greater than chance number of coincidences could be counted.

Results

Details are in Table III. These show that the similarity is restricted almost entirely to the first digit.

(*iii*) *Comparing intrusions with correct responses*

A similar procedure was employed to that used to compare intrusions with omissions. This showed that the mean number of common digits per subject was 1·86, whereas the number expected by chance would be 1·84. Eleven subjects had more than would be expected by chance. Twelve had fewer. Clearly there is no similarity between intrusions and items correctly recalled.

TABLE III

THE COMPARISON OF INTRUSION AND OMISSIONS IN FREE RECALL SHOWING THE MEAN PERCENTAGE OF COMMON DIGITS AND THE NUMBER OF SUBJECTS WITH MORE COINCIDENCES THAN WOULD BE EXPECTED BY CHANCE (S)

Kind of coincidence (*see text*)	*Original study* ($n = 49$)			*Replication* ($n = 62$)		
	Observed	*Chance*	*S*	*Observed*	*Chance*	*S*
a (common first digits)	16·7	10·5	37*	15·0	10·7	48*
b (common second digits)	9·9	9·4	24	11·4	9·2	34
c	10·3	9·6	19	10·8	9·9	39
d	11·8	11·0	22	9·9	11·3	25

* $p < 0·001$

PART V. ORDER OF RECALL

Casual inspection of the raw data of Part I indicated that subjects were inclined to write the numbers they recalled in ascending order. Accordingly, it was decided to analyse the replication data systematically to check this observation. It was for this reason that the order of presentation had been constrained to exclude all triplets in ascending or descending order. Any ascending triplets present were evidence of ordering imposed by the subjects.

For each subject the mean number of ascending triplets was 3·27 while the mean number of descending triplets was 1·32. Of the 92 subjects (one had to be excluded since the order in which he had written out his recall could not be determined) 59 showed more ascending than descending triplets thus indicating a significant tendency ($p < 0·01$) in the expected direction.

PART VI. PROXIMITY

The evidence so far indicates that when subjects fail to remember both digits of a number correctly they show a significant tendency to retain the first one. This could be because the subjects code the numbers by their position along a scale stretching from 10 to 99 so that an error in the second digit arises through inaccuracy introduced at some stage during storage and retrieval.

In an attempt to test this hypothesis, recognition tests were constructed having distractors which were either close to, or relatively distant from, the stimuli. The argument was that if subjects remember the position of each stimulus number along a scale then they will find distractors close to the stimuli more difficult to discriminate from them than more distant distractors.

Method

A list of 15 numbers was chosen, together with distractors which had the same first digit but varied in proximity. The material is shown in Table IV. As before, multiples of 10 and 11 were excluded.

TABLE IV

MATERIAL FOR THE PROXIMITY EXPERIMENT

Stimuli ..	12	19	23	35	36	41	48	54	59	67	71	79	84	85	93
Near distractors ..	13	18	24	34	37	42	49	53	58	68	72	78	83	86	94
Distant distractors	15	16	27	31	39	45	46	51	55	62	75	76	81	89	98

Immediate memory was tested. Details of presentation and testing were the same as in the recognition experiments reported in Part II. Subjects were tested in three batches of approximately 20. A fresh order of presentation was used for each batch.

Subjects. Sixty-one young enlisted men not previously tested.

Results

The mean number of stimuli correctly identified by each subject was 10·53 when "near" distractors were employed, and 10·84 when "distant" ones were used. The very small difference in the expected direction is not statistically significant ($t = 0{\cdot}7$ with 59 *d.f.*, $p > 0{\cdot}25$). This experiment indicates, therefore, that the proximity of distractors does not appreciably increase their effectiveness.

Summary of Main Experimental Findings

(1) Intrusions in free recall are non-random.

(2) When common intrusions are used as distractors in recognition tests the test is relatively difficult, whereas the use of rare intrusions as distractors makes the test easy.

(3) Intrusions tend to have the same first-place digit as omissions.

(4) The use of ease of learning for free recall as an index of response availability led to rejection of the hypothesis that response bias influences intrusion frequency.

(5) Intrusions do not resemble items correctly recalled.

(6) The proximity in the number series of distractors to correct items does not affect the difficulty of a recognition test.

(7) In free recall subjects tend to record their responses in ascending order.

General Discussion

The observation that free recall intrusions are not random but show some similarity to omissions is significant for several reasons. It shows that this material is not learned in an all-or-none fashion, since the intrusions are partially correct responses. Another way of describing the findings is to state that the errors contain information. This means that a count of the number of completely correct responses will provide a conservative measure of the amount of information retained. It also means that a simple application of information theory to the measurement of retention as illustrated by Davis *et al.* (1961) is inappropriate.

The demonstration that the difficulty of recognition tests can be manipulated by the systematic selection of the distractors provides another illustration of the inadequacy of the simple information measure. This assumes that errors are equiprobable and that distractors are therefore equally effective so that the difficulty of a test depends solely upon their numerosity. It is possible that only a limited number of alternatives are really effective as distractors and that the effect of increasing the number of distractors affects performance through increasing by chance the number of effective distractors employed.

An analogous position has been reached in the study of immediate memory for short sequences of letters. Conrad (1964*a*) has elegantly demonstrated that performance is governed by the confusibility between items in the population of stimuli employed rather than the size of that population. Elsewhere (Conrad 1964*b*) he has argued that the effect of vocabulary size, which has been used to support the notion that the amount of information per item determines difficulty, could be an

artefact, since as vocabulary size increases so could the number of acoustically similar items within the set.

The correspondence between recall and recognition raises the issue of the extent to which common underlying processes are involved. It is probable that recall involves the dual process of (i) making the traces of as many category members as possible available for inspection by some internal scrutinizer, and (ii) deciding at the time of inspection whether or not a particular number belongs to the set of stimuli which has been exposed. The evidence that in free recall subjects tend to write out their responses in ascending order suggests that they scan from 9 to 99, and since all two-digit numbers are well known to all the subjects difficulty would only arise in the recognition stage. The observed similarity of recognition and recall performance levels when all possible two-digit numbers are included in the recognition test is therefore to be expected for this material.

The similarity between intrusions and omissions was found to be restricted to the first-place digit. When unable to give the correct response subjects tended to give some alternative response which had the same first-place digit. Wickelgren has made a similar analysis of errors in the immediate recall of consonant-vowel digrams (Wickelgren, 1965). With this material there was no tendency favouring recall of the first member of each pair, instead the consonant was favoured regardless of its position. The order effect is, therefore, not general and it could be peculiar to numbers where in many situations the first place is more important than the second.

We are indebted to Dr. C. H. Brock and Mrs. A. Hull for assistance with Data analysis and to the Ministry of Defence Depts. A and N for providing subjects and testing facilities.

References

Brown, J. D. (1959). Information, redundancy and decay of the memory trace. In *Mechanisation of Thought Processes. Nat. Physical Lab. Symp. No.* 10. London: HMSO.

Conrad, R. (1964*a*). Information, acoustic confusion and memory span. *Brit. J. Psychol.*, **55**, 429–32.

Conrad, R. (1964*b*). Acoustic confusions in immediate memory. *Brit. J. Psychol.*, **55**, 75–84.

Dale, H. C. A., and Baddeley, A. D. (1962). Alternatives in testing recognition memory. *Nature*, **196**, 93–4.

Davis, R., Sutherland, N. S., and Judd, B. R. (1961). Information content in recognition and recall. *J. exp. Psychol.*, **61**, 422–9.

Stone, M. (1960). An extension of the chi-squared test for randomness. *Brit. J. stat. Psychol.*, **13**, 31–2.

Wickelgren, W. A. (1965). Similarity and intrusions in short term memory for consonant-vowel digrams. *Quart. J. exp. Psychol.*, **17**, 241–7.

Manuscript received 19*th January*, 1966.

THE EFFECT OF INTERPOLATED ACTIVITY ON A TREBLE RIGHT-LEFT ALTERNATION TASK

BY

E. A. TAYLOR, M. J. McEWEN and L. WEISKRANTZ
From the Psychological Laboratory, Cambridge

In a delayed treble alternation task the number of errors made by human subjects was found to increase when an irrelevant task was presented in the delay interval. The number of errors increased with the number of items presented in the delay interval, but neither the difficulty of the irrelevant task nor its similarity to the alternation task was found to affect the number of errors.

Introduction

It has long been known that dorsolateral lesions in the frontal cortex of the monkey can produce an impairment of tasks requiring a delayed response to a discriminatory stimulus, so that the animal cannot choose the correct stimulus more frequently than chance if the response is delayed by more than 5–10 sec. Jacobsen (1935) characterized this as an impairment of immediate memory. Such animals can, however, still learn difficult discriminations; and so, if the defect is to be in the short-term memory store, it would seem to be a specific retrieval difficulty.

Two predictions from this have been confirmed. Mishkin and Weiskrantz (1958) found that a task involving delay of reward was also impaired after frontal lesions. Weiskrantz, Mihailović and Gross (1962) found that frontal stimulation caused impairment in learning simple problems for which few learning trials were necessary, but were not impaired in the learning of difficult problems for which many trials were necessary: this would follow if the retrieval difficulty was specific to the short-term store and so would not show up in the difficult problems for which the long-term store was involved.

Other studies, however, have cast doubt on the retrieval hypothesis. It was found by Oxbury and Weiskrantz (1962) that animals with frontal lesions were better than normal animals in the development of learning set, as indexed by improvement in performance over a series of problems for each of which three trials were given regardless of the result. It was therefore suggested that the difficulty is not in retrieval but in the registering of information: perhaps frontal monkeys select stimuli differently so that there is what might be called an alteration of attention. There are indeed indications that monkeys with frontal lesions do respond differently to complex stimulus situations; they take shorter glimpses than normals (Lindsley, Weiskrantz and Mingay, 1964) and they sample from the stimulus alternatives in a more chaotic manner (Weiskrantz and Mingay, unpublished) particularly as the number of choices is increased. Buffery (1964) found that animals with frontal lesions departed radically from the controls as the number of alternatives in a discrimination increased. This effect increased as the novelty of the negative stimuli increased. He also varied the frequency with which four stimuli were presented in a delayed matching from sample task, and found that the operated animals responded better than controls to the infrequent stimuli and worse to the frequent.

This chaotic stimulus sampling may be analogous in some respects with that seen in human manic patients. Meprobamate is given as a tranquilliser to these patients; the effect of meprobamate on animals with frontal lesions was investigated by

Weiskrantz, Gross and Baltzer (1965). The drug had no effect on delayed response, but did have a "sensory inhibitory" effect in normal animals. The drug significantly improved the performance of operated animals on a delayed response task. This would be expected if it is countering the "sensory disinhibition" of the frontal animals.

There is also some evidence that frontal animals perform better when their delay interval is spent in darkness than when it is in normal illumination: this also is compatible with the theory outlined above (Malmo, 1942; Orbach and Fischer, 1959).

For this theory to hold it is necessary that paying attention to irrelevant stimuli in the delay interval should impair performance on a delayed response task. This effect has been shown in normal humans in experiments where the delayed response is recall of a list of digits, e.g. by Brown (1958) and by Conrad (1960); the present experiment was designed to study the effect in a situation more comparable to that of the animal experiments. Preliminary studies suggested that the effect could be shown in a delayed treble alternation task with a constant delay interval of 8 sec.

Method and Apparatus

The main task was a triple right-left alternation; two piles of playing cards were placed face down in front of the subject and he was required to take up the cards, one at a time, in the order:

one from the right heap
one from the right heap
one from the right heap
one from the left heap
one from the left heap
one from the left heap
one from the right heap
one from the right heap
one from the right heap
etc.

with a delay of 8 sec. between each card.

Each subject had 30 cards to turn up on each run; the score was the number of errors made. Knowledge of results was desirable since otherwise great difficulties of strategy were presented. The cards were therefore arranged so that when the subject turned up a correct card it was red; when he was wrong the card was black. This was accomplished by making the two piles of 30 cards each, arranged in the following way from the top of the pile down:

Right pile: red, red, red, black, black, black, red, red, red, etc.
Left pile: black, black, black, red, red, red, black, black, black, etc.

When the subject turned up a card from one pile, the experimenter took a card from the opposite pile. In this way there was an objective sequence with which the subject had to keep in step. This had the advantage of simplifying scoring but the disadvantage that some element of reasoning as well as of memory was involved in getting back into sequence after making a mistake. Subjects did not, however, have any difficulty in this. Another advantage of the method was that it preserved a record of the run which could be analysed afterwards and also retained information about the order of the mistakes (which were distributed evenly through the run and did not cluster in any one position).

The experimental design varied the number and difficulty of items presented in the delay interval, which was maintained constant at 8 sec. There were four conditions of number and three of difficulty: in addition each subject had a control condition in which no items were presented during the delay interval. The "irrelevant" task involved a response to the presentation of numbers in the window of a memory drum. In a window, measuring 1 in. wide and $\frac{5}{8}$ in. high, on the top of a memory drum, were presented the numbers "1" and "2," the digits being approximately $\frac{3}{8}$ in. high. Either the "1" was on the right and the "2" on the left or the "1" was on the left and the "2" on the right. Screwed to the top of the memory drum was a writing platform. A lined strip of paper

passed by the right hand side of the window and the subject had to tick off the position of the "1" on this strip of paper. Thus, he put a tick on the left when the "1" was presented on the left, and on the right when the "1" was presented on the right. Also on the writing platform, but to the left of the window, were two piles of cards used in the main task described above; the right hand pile was approximately 2 in. to the left of the window. For left-handed subjects the arrangement was reversed and the strip of paper passed by the left-hand side of the window while the piles of cards were to the right of the window.

In one condition of difficulty (A) of the items presented in the delay interval, the "1" was always presented on the left side of the window and so conveyed no information. In condition B, the "1" was presented either on the left or on the right in random order, thus conveying one bit of information. In condition C the "1" was presented on the left and on the right alternately; again, this strictly conveyed no information but, inasmuch as it was an alternation, was similar to the main task and so might be expected on a reactive inhibition theory to interfere more than A with the main task.

Condition D was a control in which no figures were presented and no ticking required.

Each of these conditions was presented at different speeds:

I. Two items were presented in the interval, one every 4 sec.
II. Four items were presented in the interval, one every 2 sec.
III. Eight items were presented in the interval, one every 1 sec.
IV. Two items were presented, one every 1 sec. at the beginning of the period, followed by a period of 6 sec. in which no items were presented.

After the 8-sec. interval, a black area appeared on the memory drum: this lasted for 4 sec. and was the cue for the subject to pick up one card, after which period items were again presented. This period of 4 sec. appeared optimal from the preliminary studies, allowing all subjects time to make their choice without undue dawdling. The memory drum was checked against an electronic timer and found to be a reliable instrument.

Thirty-two subjects were used, all undergraduate students at the University of Cambridge. Each had one run in each of the conditions A, B, C and D, the order in which they were given being varied between subjects according to a Latin square design. Each subject therefore had four runs, all at one speed; different subjects were assigned at random to one of the four speed conditions, I, II, III or IV.

Each run lasted 6 min.; there was a pause of 3 min. between each run. Standardized instructions were read to each subject and a short demonstration then given by the experimenters. In these instructions the subject was given no indication that either of the two tasks was more important than the other. The instructions included a request not to "cheat on some part of the experiment by giving yourself clues, for instance, by using the position of your hand between cards to indicate the right side, or by repeating to yourself the position of the next card."

Results

The results for individual subjects are shown in Table I. The totals for the different conditions are shown in Table II. Only errors made on the main task are considered; each cell represents the results of eight subjects added together. Each figure is therefore out of a possible total of 240, except the column totals and row totals, which are out of 960.

To determine whether activity presented in the delay interval affected the main task, each subject's score on the control, D condition was compared with his averaged score on the three other conditions. The number of times D was greater than $(A+B+C)/3$ was then compared with the number of times it was smaller, and a signs test indicated a probability associated with the null hypothesis (that there was no difference between D and $(A+B+C)/3$ of less than 0·00001 (on a one-tailed test). The result is therefore very significant. All subjects but three showed a change in the predicted direction. Even when scores on the D condition are compared with those on the A condition, D is found to be significantly smaller than A, at below the 0·002 level on a one-tailed test.

TABLE I

Subject	Condition	Task A (unilateral)		Task B (random)		Task C (alternate)		Task D (control)
		X	Y	X	Y	X	Y	X
1	I (2 spaced)	0	0	0	0	0	0	0
2	II (4 spaced)	2	2	2	2	2	2	0
3	III (8 spaced)	8	0	3	9	3	6	2
4	IV (2 massed at start)	2	0	0	0	0	0	0
5	I	2	0	1	0	1	0	1
6	II	1	0	4	5	5	0	0
7	III	1	0	2	0	2	0	1
8	IV	0	0	0	0	1	0	1
9	I	0	0	1	0	0	0	0
10	II	3	0	2	1	1	3	0
11	III	4	0	7	6	7	1	1
12	IV	1	0	0	1	1	0	0
13	I	3	0	2	0	1	0	1
14	II	1	0	1	6	1	1	0
15	III	5	0	2	1	9	0	1
16	IV	3	0	0	0	4	0	0
17	I	0	0	6	0	0	0	3
18	II	4	0	3	2	4	0	3
19	III	2	0	6	35	3	0	0
20	IV	2	6	1	1	1	0	0
21	I	0	0	1	0	1	0	0
22	II	1	0	2	1	5	1	0
23	III	7	0	4	12	6	4	0
24	IV	1	0	2	0	1	0	1
25	I	1	0	3	1	3	0	1
26	II	1	0	0	1	1	0	1
27	III	0	0	1	30	1	3	0
28	IV	0	0	4	7	8	2	0
29	I	5	0	2	0	2	0	0
30	II	1	0	0	3	1	0	0
31	III	3	1	12	25	4	0	0
32	IV	0	0	2	0	0	0	0

Columns marked X give the number of errors made on the main task; each figure is out of a possible maximum of 30.

Columns marked Y give the number of errors made on the irrelevant task; at speed I each figure is out of a possible maximum of 58, at speed II out of 116, at speed III out of 232 and at speed IV out of 58.

The signs test also gives the following two-tailed probabilities associated with the null hypotheses that there is no difference between:

A and B: $p < 0{\cdot}9$.
B and C: $p < 0{\cdot}64$.
C and A: $p < 0{\cdot}38$.

It is therefore impossible to reject the null hypotheses.

A Friedman two-way analysis of variance was carried out on the data in Table II. This indicated a probability of 0·653 that the three samples A, B and C had been drawn from the same population: it must therefore be concluded that this experiment shows no significant differences between the different conditions of difficulty. Condition D was excluded from the Friedman analysis of variance described above, since no items at all were presented in this condition.

TABLE II

Conditions	*Task A*	*Task B*	*Task C*	*Task D*	*Summary*	*A + B + C Total errors for each condition*
I	11	16	8	6	41	35
II	14	14	20	4	52	48
III	30	37	35	5	107	102
IV	9	9	16	2	36	34
Total errors for each task	64	76	79	17		

The Friedman test was then applied to the four-speed conditions I, II, III and IV. This indicated a probability of 0·054 that there was no difference among the conditions, which is just short of significance. This, however, was ascribed to the effect of including condition IV in the analysis; since, like I, it had two items presented in the delay interval, it may be regarded as being the same as I. When conditions IV and I were combined and averaged, analysis of variance indicated that the three conditions of speed I, II and III were significantly different, at the 1 per cent. level.

A Mann-Whitney U test was therefore carried out on the data of Table I. This indicated the following two-tailed probabilities associated with the null hypotheses that there was no difference between:

I and II : $p < 0{\cdot}234$.
I and III : $p < 0{\cdot}014$.
I and IV : $p < 0{\cdot}798$.
II and III : $p < 0{\cdot}028$.
II and IV : $p < 0{\cdot}194$.
III and IV : $p < 0{\cdot}006$.

It will be noted that condition III impaired recall significantly more than any of the other conditions (at 3 per cent. level of confidence). There were no other significant differences among the different conditions of speed.

A parametric analysis of variance confirmed all these findings and moreover indicated that there was no significant interaction between the conditions of speed and of difficulty. The data were satisfactory from the point of view of homogeneity of variances, but it cannot be assumed that the other requirements of the analysis of variance were fulfilled.

Discussion

(1) *Differences between tasks A, B, C and control (D)*

The differences between A, B, C and the control task D were highly significant, showing that interpolated activity in the delay interval did impair a delayed

alternation task. Although this task is essentially non-verbal, unlike the more usual experiments involving recall of a list of digits or letters, verbal rehearsal was possible. As has already been mentioned, subjects were told not to rehearse verbally, but it is uncertain how far they followed these instructions. Like the subjects of Pillsbury and Sylvester (1940) they reported that it was very hard not to, and they were uncertain as to how far they had obeyed the instructions. Other experiments have indicated that instructions not to rehearse may not have clear effects: McLane and Hoag (1943) found in a class-room situation that there was no decline in recall with time when instructions not to rehearse were given. It might be interesting to attempt a similar experiment to the present one using relatively unrehearsable material.

(2) *Differences between the difficult conditions A, B and C*

There were no significant differences between the conditions of difficulty A, B and C. This is in contrast with the suggestion by Pillsbury and Sylvester (1940) that the amount of the decrement of recall of pictures, nonsense syllables or words was more closely related to the overall difficulty of the interpolated task than to its similarity to the recall material. The absence of any difference between C and A or B further suggests that the similarity of the interpolated to the recall material is not an important variable. This is consistent with Brown's (1958) finding that in recall of digits similarity of the intervening task did not produce significant differences in forgetting. In different situations, however, the effects of similarity have been shown. Baddeley (1964) compared the effects of acoustic and semantic similarities of the different words in the list to be recalled. A small but significant impairment of recall was found in the semantically similar lists, and a very much larger impairment of recall in the acoustically similar lists. However, he interprets this finding in terms of a strong acoustic factor in immediate memory, rather than in terms of a general effect of similarity in impairing recall.

One possible explanation of our findings might be that different numbers of mistakes were made on the interpolated task so that the information transmitted did not vary greatly over the different difficulty conditions. In fact it can be seen that in the fastest speed condition, III, every subject, except one, made more mistakes on the interpolated task B than on the tasks A and C, and that the exceptional subject, 7, made no mistakes on any of the interpolated tasks. Even when as many as 35 mistakes were made, however, the information transmission, strictly considered, was still high at 232–35 = 197 bits/run. Certainly subjective reports stressed that condition B was more difficult than the rest. It is also possible that the failure to find any differences between conditions was due to there being an insufficient gradient of difficulty between them to show up on our measure. In particular, there was a short time in which the subjects could appreciate the redundancy of A and C.

If the lack of differences between conditions is a real effect, it might be due to a second factor offsetting the effects of difficulty. It might be that subjects are more aroused in the more difficult condition and so make fewer mistakes. They did complain that condition A was monotonous. If this is so, however, it is surprising that the same effect was not seen when the speed of presentation of items was increased. It should also be noted that the reason why conditions A and C transmitted no information was that subjects remembered the past sequence; no such memory would be involved in condition B. It is possible that the extent to which the irrelevant task involves the use of the memory store is an important determinant of the extent to which the task impairs recall.

(3) *Differences between speed conditions I, II, III, IV*

Since the interval was kept constant there was a confounding in this experiment of the effects of speed of presentation and number of items.

It is clear that condition III in which eight items were presented in the 8-sec. interval gave significantly more mistakes than did any of the other conditions. It therefore appears that varying the number of items presented in the interval does have an effect on the main task; the failure to find any significant difference between conditions I and II is presumably due to the insensitivity of the method. The former finding can be taken as confirming one of the predictions of decay theory, i.e. that the detrimental effects should vary directly with the extent of reduction in rehearsal time, if it is assumed that some rehearsal was taking place in our experiment in spite of the contrary instructions. This is consistent with the findings of Brown (1958) in a situation where number of items and length of interval were confounded. Conrad (1960) found that when a longer series of digits than in Brown's experiment was used, the single interpolated response of saying "zero" produced a large amount of interference. Postman (1964) interprets this to mean that it is not the length of the interval but the fact of interpolation *per se* which is responsible for the retention decrement in this situation. A later study by Pollack (1963) shows that if the number of interference digits per second is kept constant and longer intervals are produced by adding more digits, then there is a positive correlation between interference duration and decrement in recall scores. If, however, the number of interference digits is kept constant and longer intervals are produced by spacing out the digits, then there is a negative relationship between interference duration and decrement in recall scores. It therefore appears, as is consistent with our results, that the number of items presented in the interval is crucial. In any case, we would expect our results to be more comparable with those of Brown than those of Conrad, inasmuch as the main task is easier than Conrad's.

The lack of significant difference between conditions I and IV indicates that it makes no difference whether the two digits are massed at the beginning of the delay period or evenly distributed throughout it. This suggests that it is the number and not the rate of presentation of items that is important.

References

Baddeley, A. D. (1964). Semantic and acoustic similarity in short term memory. *Nature*, **204**, 1116.

Brown, J. (1958). The decay theory of immediate memory. *Quart. J. exp. Psychol.*, **10**, 12.

Buffery, A. W. H. (1964). Ph.D. thesis in the University of Cambridge.

Conrad, R. (1960). Very brief delay of immediate recall. *Quart. J. exp. Psychol.*, **12**, 45–7.

Jacobsen, C. F. (1935). Functions of the frontal association area in primates. *Arch. Neurol. Psychiat.*, **33**, 558–69.

Lindsley, D. F., Weiskrantz, L., and Mingay, R. (1964). Differentiation of frontal, inferotemporal and normal monkeys in a visual exploratory situation. *Anim. Behav.*, **12**, 525–30.

McLane, A. S., and Hoag, J. E. (1943). The curve of forgetting in the first three minutes. *Amer. J. Psychol.*, **56**, 105–10.

Malmo, R. B. (1942). Interference factors in delayed response in monkeys after removal of frontal lobes. *J. Neurophysiol.*, **5**, 295–308.

Mishkin, M., and Weiskrantz, L. (1958). Effects of delaying reward on visual discrimination performance in monkeys with frontal lesions. *J. comp. physiol. Psychol.*, **51**, 276–81.

Orbach, J., and Fischer, G. J. (1959). Bilateral resections of frontal granular cortex. *Arch. Neurol.*, **1**, 78–85.

OXBURY, J., and WEISKRANTZ, L. (1962). Effects of frontal lesions on object discrimination learning in monkeys. *Nature*, **195**, 310–1.

PILLSBURY, W. B., and SYLVESTER, A. (1940). Retroactive and proactive inhibition in immediate memory. *J. exp. Psychol.*, **27**, 532–45.

POLLACK, I. (1963). Interference rehearsal and short term retention of digits. *Canad. J. Psychol.*, **17**, 380.

POSTMAN, L. (1964). Short-term memory and incidental learning. In MELTON, A. W. (Ed.), *Categories of Human Learning*. New York: Academic Press.

WEISKRANTZ, L., GROSS, C. G., and BALTZER, V. (1965). The beneficial effects of meprobamate on delayed response performance in the frontal monkey. *Quart. J. exp. Psychol.*, **17**, 118–24.

WEISKRANTZ, L., MIHAILOVIĆ, L., and GROSS, C. G. (1962). Effects of stimulation of frontal cortex and hippocampus on the behaviour in the monkey. *Brain*, **85**, 487–504.

Manuscript received 27th January, 1966.

EXTINCTION OF A RUNNING RESPONSE AS A FUNCTION OF REINFORCEMENT MAGNITUDE

BY

R. N. WILTON

From the Department of Psychology, University of Exeter

Four groups of rats were trained on different sucrose solutions in a straight runway. Terminal running speed was a monotonic function of reinforcement magnitude. After training each group was subdivided, one subgroup being extinguished under spaced, the other under massed conditions. In spaced extinction the animals trained on non extreme reward magnitudes showed most resistance to extinction. It was concluded that resistance to extinction is an inverted U-shaped function of reinforcement magnitude found in training. The massed extinction trials were conducted with a very short inter-trial interval. The animals showed an immediate drop in running speed followed by a gradual recovery and a subsequent decline. The number of trials taken to reach the peak recovery speed was a function of reinforcement magnitude found in training. Results on both massed and spaced extinction trials were interpreted in terms of the facilitatory and inhibitory effects of momentary and conditioned frustration.

Introduction

The relationship between magnitude of reinforcement and resistance to extinction is unclear. Experiments in the straight runway (Hulse, 1958; Armus, 1959; Wagner, 1961) demonstrate faster extinction for rats trained on the larger of two rewards. Hill and Spear (1962) pointed out that these results were at variance with those Skinner box studies (e.g. Guttman, 1953) which show that resistance to extinction is a positive monotonic function of reinforcement magnitude. In their own runway study, resistance was greater following training with a larger reward on both massed and spaced extinction trials. The results parallel those of the Skinner box but conflict with the other runway studies.

All the runway experiments used only two reinforcement values. This points to the possibility that some disparity in results may be due to the sampling of different segments of a curve which shows a non-monotonic relationship between reward magnitude used in training and performance in extinction. The present work therefore reports the results of massed and spaced extinction after training on one of four reward magnitudes in a straight runway.

Amsel (1958) has demonstrated that extinction trials generate an increase in motivation which increases the vigour of responses immediately following non-reward. The effects of this temporary increase in motivation could be different in the chained trials of the Skinner box and the spaced trials of the runway, again accounting for some differences in results. Accordingly the massed trials were conducted with a minimal inter-trial interval, allowing the immediate effects of non-reward to influence behaviour on subsequent trials. In this respect massed extinction more nearly corresponded to the Skinner box situation.

Method

Subjects. These were 48 experimentally naïve male hooded rats, aged 10–12 weeks at the start of the experiment.

Apparatus. This consisted of a runway 6 ft. long and 5 in. high with flat grey wooden sides and a wire mesh floor and roof. The start box section, 12 in. long, was painted black. Latency and running times were combined and recorded on a timer, operated by a manually lifted guillotine door at the start box and by a photocell 2 in. from an aluminium tray which opened into the end of the runway. A $\frac{1}{2}$ in. diameter hole in the floor of the $1\frac{1}{2} \times 2 \times 2\frac{1}{2}$ in. tray enabled the rat to obtain sugar solution from a small fountain produced by a pump. This was activated by an interruption of the photobeam.

Reinforcement. Four groups of rats were trained on different reinforcement magnitudes. Group 1 was given 5 sec. access to a 6 per cent. sucrose solution. Group 2, 50 sec. of 6 per cent. solution. Group 3, 5 sec. of 50 per cent. solution. Group 4, 50 sec. of 50 per cent. solution. Solutions were made up by weight of sucrose to volume of water.

Pretraining. The animals were individually handled for 5 min. per day on the first three days of pretraining. On the fourth and fifth days each rat was allowed to explore the alley, the start door and pump having been removed. A 22¾ hr. deprivation schedule began on the fourth day; subjects received unlimited food for 1¼ hr. per day. On the sixth and seventh days each rat was twice placed in the final section of the alley and allowed approximately 15 sec. access to its appropriate concentration of reinforcement.

Training. A training trial began by placing an animal in the start box facing the door. The door was lifted 2–3 sec. later, allowing the animal to run the alley. Interruption of the photobeam stopped the timer and activated the pump. Each animal was allowed access to its appropriate reinforcement. The pump then automatically switched off and the rat was removed 2–3 sec. later. Each rat was given a total of 48 trials, three on the first day and five thereafter. Total training time was 10 days. The minimal inter-trial interval was 40 min. The rats were fed ½ hr. after the last trial. Deprivation conditions were equated for all groups and ranged from 16–22 hr. for each member of each group on each day of training.

Extinction. Each reinforcement group was split into two. Four groups (1M, 2M, 3M, 4M) were extinguished under massed and four (1S, 2S, 3S, 4S) under spaced conditions.

The pump, which gave a slight hum during operation, was placed in a position inaccessible to the rats about 1 in. from the reinforcement tray. The experiment was concerned with the relationship between the groups on either one of two extinction procedures. Corresponding groups in massed and spaced procedures were not to be compared. Hence it did not seem inappropriate to delay spaced extinction until massed extinction was complete.

Massed. This took place in the two days which followed training. Half the members of each group were completely extinguished on the first day, the other half on the following day. The animals were run in such an order that deprivation conditions were equated for all groups. A trial proceeded as follows: A rat was placed in the start box; 2–3 sec. later the door was raised allowing the rat to run the alley and break the photobeam. The pump operated for 5 sec. The rat was removed 2–3 sec. later and *immediately* returned to the start box. This procedure was repeated for a total of 50 trials.

Spaced. The spaced animals were rested in the two days following training. They were then given 50 extinction trials; five per day for 10 days. Minimum inter-trial interval was 40 min. A trial proceeded as follows: A rat was placed in the start box and 2–3 sec. later the door was raised. The pump operated for 5 sec. following a break in the photobeam. The rat was removed 2–3 sec. later and returned to his living cage.

If on any trial an animal's running time exceeded 90 sec. he was removed; animals on massed extinction were returned to the start box, those on spaced extinction to the living cage. A time of 90 sec. was recorded.

Results

Training. The approximate mean volume of solution consumed by each animal on each trial was as follows: Group 1, 0·2 c.c.; Group 2, 1·0 c.c.; Group 3, 0·2 c.c.; Group 4, 1·0 c.c. Those rats given a longer exposure time thus drank considerably more fluid.

Group mean running speeds were ordered in the sequence 1, 2, 3, 4 (Fig. 1). Initial differences were maintained as response speeds levelled off. A Mann Whitney analysis (from Siegel, 1956) of mean speeds on the terminal trials 44–48 gives a $p < 0{\cdot}01$ for any two groups except 3 and 4. The results confirm those of Kraeling (1961) which show that increases in volume have similar but smaller effects on performance than increases in concentration.

Spaced extinction. As the experimental groups began extinction at markedly different levels of speed the rate of extinction was measured according to a method

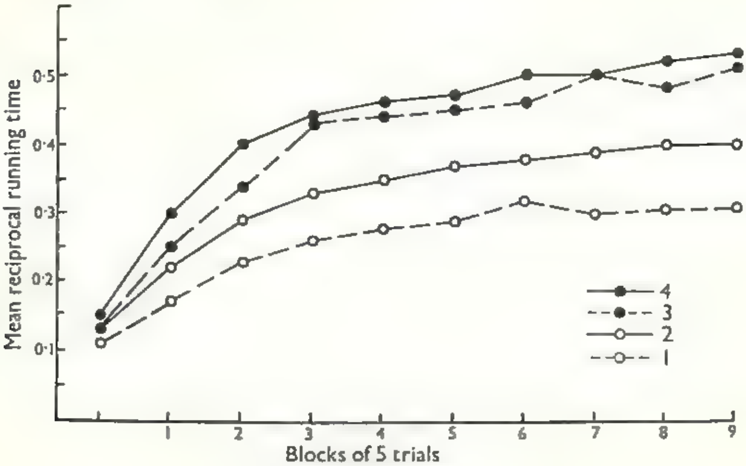


FIGURE 1
Running speed in spaced training.

suggested by Wagner (1961). The speed of each rat over blocks of five extinction trials was expressed as a proportion of the mean speed attained on its last five training trials. The mean proportions from extinction trials 1–30 was taken as an index of resistance to extinction. This measure seemed most compatible with the 32, 19, and 20 trial measures adopted by Wagner, Hulse, and Hill and Spear.

Non-parametric analysis of variance (Kruskall-Wallis) gave a $p < 0{\cdot}05$ for differences between the four groups. This justified the use of the Mann Whitney test for analysis of difference between any two groups. Table I shows the mean proportional speeds for the groups and their associated probabilities of difference. Groups 2S and 3S are most resistant to extinction. One way of summarizing the results is to simply state that most resistance to extinction is obtained from animals trained on a combination of weak concentration and large volume or strong concentration and small volume. A strong concentration and large volume or a weak concentration and small volume is followed by low resistance to extinction.

However, since differences in concentration and volume have generally been shown to have similar effects on performance measures (Kraeling, 1961; Collier and Myers, 1961) it seems more appropriate to place the groups on a continuum of reinforcement magnitude. Each group can be positioned according to its terminal training speed; this is known to be a monotonic function of reinforcement magnitude (Logan, 1960; Kraeling, 1961). Extinction rate is now seen to be an inverted U shaped function of reinforcement magnitude. On this basis it is possible to explain

TABLE I

Probability of Group Differences in Spaced Extinction and Mean Rate of Extinction (in brackets)

Groups	*p*
1S (2·26) and 2S (3·15) ..	0·047
1S and 3S (2·74)	0·021
1S and 4S (1·97)	0·155
2S and 3S	0·294
2S and 4S	0·013
3S and 4S	0·013

the disparity in results of previous spaced studies. Hulse (1958) and Wagner (1961), who both used 1 and 0·08 gm. pellets, found fastest extinction following the larger reward; Hill and Spear (1962), using 0·18 and 0·045 gm. pellets, found fastest extinction following the smaller reward. A comparison with the present work suggests that these studies sampled different portions of the non-monotonic extinction curve, and that Hill and Spear's largest 0·18 and Hulse's smallest 0·08 groups are analogous to the 2S or 3S groups. Their apparently contradictory results are therefore to be expected.

Figure 2 shows that *absolute* speed in later extinction is also an inverted U shaped function of reinforcement magnitude. Analysis of mean speeds on trials 26–45 gives 2S as the significantly fastest group. ($p < 0{\cdot}05$ when 2S is compared with any other group). As both later extinction speed and extinction rate are non-monotonic functions of reward magnitude it seems probable that they are determined by common variables. On the other hand group 3S is significantly different on rate but not on terminal speed, suggesting that the measures are not completely equivalent. However, the early extinction curve for 3S looks atypical and no conclusion may be drawn with confidence.

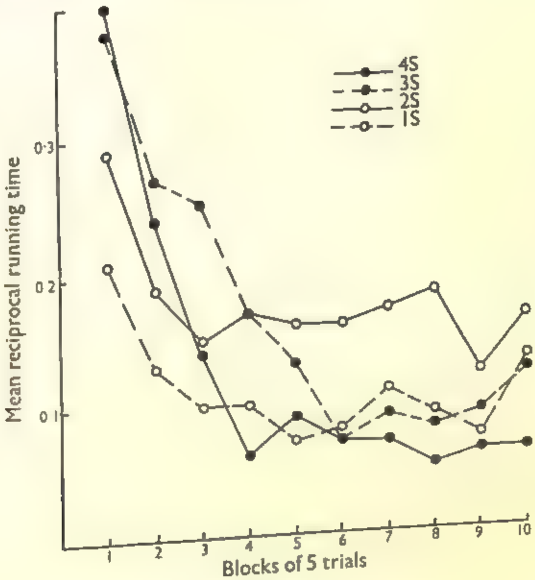


FIGURE 2

Running speed in spaced extinction.

Massed extinction. The massed results demonstrate an effect only incidentally reported in the previous literature (Stanley and Rowe, 1954; Young and Shuford, 1955; Weinstock, 1958; Lawrence and Festinger, 1962). The speed of each group shows an initial decline. This is followed by a gradual increase to a peak and then another final decline. A trend analysis of speed increase from trial 10 to the peak gives the following results: Group 1, trials 10–50, $p = 0{\cdot}0003$; Group 2, trials 10–45, $p = 0{\cdot}007$; Group 3, trials 10–40, $p = 0{\cdot}19$; Group 4, trials 10–25, $p = 0{\cdot}05$. A further non-parametric analysis (Jonckheere, 1954) shows that animals trained on larger rewards tend to reach their peak extinction speeds on earlier trials; $p = 0{\cdot}006$ (Fig. 3). Thus the position of the extinction "peak" is a function of reinforcement magnitude.

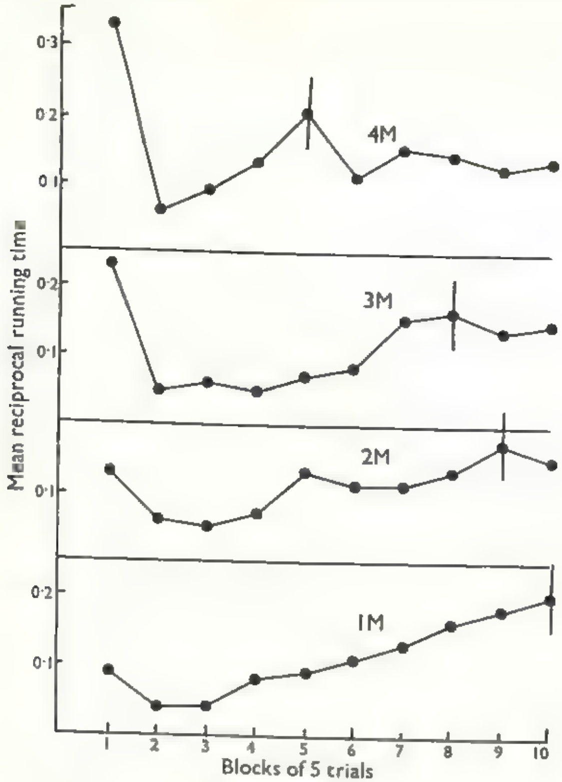


FIGURE 3
Running speed in massed extinction.
Position of peaks is marked.

Discussion

An explanation of extinction might be given in terms of frustration theory (Amsel, 1958). This asserts that non-reward at a goal generates an aversive motivational state termed frustration. The standard apparatus for measuring this motivational consequence of non-reward is the double runway. A rat is trained to run down a first runway to a goal box (*G*1) for food. An exit door is then lifted and the rat runs down a second runway for food in a second goal box (*G*2). When reinforcement is omitted in *G*1 running speed in the second runway is considerably increased. This increase in vigour, called the frustration effect (*FE*), is said to be due to the evocation of a frustration response (*RF*) in the first goal box.

Amsel has argued that the aversive properties of frustration play a part in extinction. A frustration response (*RF*) is elicited on each non-reinforced trial. Alley stimuli then come to elicit antedating frustration responses (r_f) and their characteristic stimuli (s_f) by stimulus generalization and by classical conditioning from the goal box. Responses evoked by s_f in the alley are assumed to be incompatible with running to the goal. With successive non-reinforced trials the competing responses evoked by s_f disrupt approach responses evoked by the antedating goal response r_g–s_g, running time increases, and extinction occurs. The analysis suggests that an approach tendency (r_g) and an avoidance tendency (r_f) jointly determine extinction speed.

All the relevant studies (see Logan, 1960) demonstrate that start box latency and running speed are negatively accelerated increasing functions of reinforcement magnitude. As these measures are directly determined by the antedating response

r_g–s_g (approach tendency) then this may be plotted as a function of similar type (Fig. 4A).

There is only one study which demonstrates a relationship between the frustration effect and reinforcement magnitude. Bower (1962) trained rats on a constant reward magnitude in the first goal box of a double runway. He then reduced this reward on certain test trials. A frustration effect was measured as an increase in speed in the second runway. This frustration effect was found to be a *linear* function of the reduction in reward magnitude found in the first goal box. It is therefore appropriate to plot the intensity of the frustration response (RF) and its antedating response (r_f) as a linear function of the reduction in reward magnitude experienced at the goal. In the present experiment all subjects were tested under similar zero reward extinction conditions but different reward magnitudes had been used in training. As in Bower's experiment this gave different reductions in reward magnitude for different groups. On this basis it seems reasonable to plot the extinction frustration response (RF, r_f) as a linear function of the reward magnitude used in training (Fig. 4A).

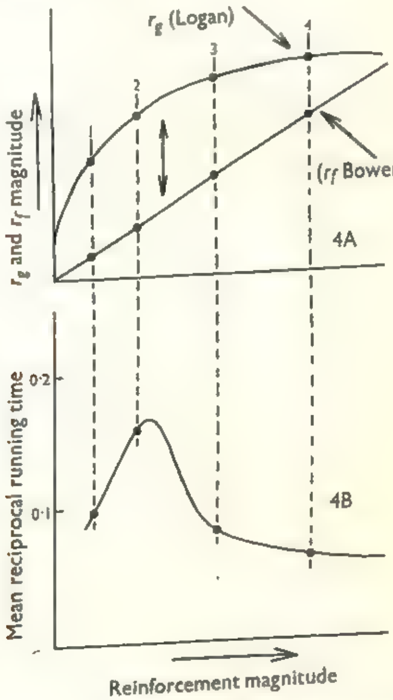


FIGURE 4

A. Empirical curves from Logan and Bower plotting magnitude of r_g and r_f as a function of reinforcement magnitude.

B. Mean running speed on trials 26–45 as a function of reinforcement magnitude.

The maximum intensity difference between points on the linear r_f and the exponential r_g curves appears at some non extreme reinforcement value. At this value the dominance of r_g over r_f will result in little competing behaviour and hence short running times in extinction. At either a lesser or greater value the intensity of r_f approaches that of r_g, competing responses become more predominant, and running time is expected to increase.

In Figure 4A the four groups are positioned on the r_g curve. This was achieved

by plotting their differences in asymptotic *training* speed on the vertical axis. The known shape of the curve (taken from Logan) and the positions of the vertical points determine the corresponding points on the horizontal axis.

In Figure 4B group mean running speed on the later extinction trials 26–45 is plotted as a function of reinforcement magnitude. Previous deduction from the r_g and r_f functions shows that extinction speed is expected to be maximal for some intermediate reinforcement magnitude. Figure 4B demonstrates that this is the case; the experimental data gives the required inverted U. The integration of these curves in Figure 4 gives strong support for a frustrative interpretation of spaced extinction. It would, however, be unwise to speculate on the data concerning extinction rate. Factors such as the growth of r_f have to be taken into account; even if this was shown to be a simple function of number of extinction trials, it would be difficult to predict the exact course of extinction for groups as small as the ones used here.

The preceding analysis cannot of course be applied to the massed results where an initial decline is followed by a peak extinction speed. Some rats ran faster over five trials at some point in massed extinction than at any time in training. This included one rat in group 1M which, after an extension of extinction trials to 140, was running faster than ever before. This implies that although the initial speed decline may be due in part to a change in stimulus conditions, produced by the shift from spaced to massed trials, simple adaptation to the new situation cannot be wholly responsible for the subsequent increase in speed. It would appear that the animals were running under added motivation during some stages of extinction. When, as in this case, a minimal inter-trial interval is used the situation becomes analogous to the double runway. Here frustration, generated by non-reward in the first goal box, increases motivation for the immediately subsequent response of running in the second runway. In massed extinction any momentary frustration produced by the absence of reward in the goal box will still be present when the animal is replaced in the start box; thus providing extra motivation for running on that trial. The fact that the position of the extinction peak is a function of reward magnitude gives weight to this interpretation. It is to be expected that those animals trained on larger reward, being more frustrated, would show motivational effects of frustration on earlier extinction trials. It should be noted that the groups (especially 4M) show a rather rapid decline from their peak extinction speeds. The reasons for this are not clear. It could be argued that conditioned frustration (r_f) plays a part in massed, just as in spaced extinction. In that case conditioned frustration, evoked by the alley stimuli, would increase progressively in the direction of the goal box. Consequently a *gradient* of frustration would inhibit approach behaviour. For those animals trained on large rewards the effects of both momentary frustration (*RF*) and the conditioned frustration (r_f) gradient would operate in the first stages of extinction. Thus the facilitatory effect of the former and the inhibitory effect of the latter would produce a successive rise and fall in running speed. Animals trained on large rewards would show this effect earliest in extinction.

Alternatively the facilitatory effects of frustration may only operate at low intensities. Higher levels of frustration might evoke competing responses which disrupt rather than enhance performance. This suggestion is, in some respects, similar to Keller and Schoenfield's (1950) interpretation of the irregular first extinction curve of the typical operant. An instance of the type of effect may be found in the Yerkes Dodson law. The manner in which anxiety influences behaviour is dependent upon its intensity. Low intensity levels facilitate ongoing behaviour, higher levels

disrupt it. Frustration and anxiety are both motivational constructs; it is quite possible that similar laws govern the action of both. Since increases in speed are found in some studies of spaced extinction (Young and Shuford, 1955; Weinstock, 1958; Lawrence and Festinger, 1962) where momentary unconditioned frustration cannot operate, this second explanation might seem the better one. As frustration grows it would first facilitate then inhibit responding, whether on massed or spaced trials. The strength of this explanation is also its weakness; some increase in speed must be predicted for all extinction procedures. The vast majority of studies show no such effect. In fact this author has found only the four mentioned here. Three of these are spaced and there are no obvious procedural similarities between these which are not shared by some other studies. Further work is required to determine the circumstances in which an increase in speed occurs in spaced extinction.

Finally it should be noted that whilst the massed results may not be directly comparable with the Skinner box studies they do demonstrate that fine temporal differences between trials can have a pronounced effect on the way in which motivation influences behaviour. It might be that momentary frustration in the highly massed extinction procedure of the Skinner Box has a facilitatory effect on a whole chain of responses. This would result in more responses in extinction for animals trained on large as compared with small rewards.

References

Amsel, A. (1958). The role of frustrative nonreward in noncontinuous reward situations. *Psychol. Bull.*, **55**, 102–19.

Armus, H. L. (1959). Effect of magnitude of reinforcement on acquisition and extinction of a running response. *J. exp. Psychol.*, **58**, 61–3.

Bower, G. H. (1962). The influence of graded reductions in reward and prior frustrating events upon the magnitude of the frustrating effect. *J. comp. physiol. Psychol.*, **55**, 582–7.

Collier, G., and Myers, L. (1961). The loci of reinforcement. *J. exp. Psychol.*, **61**, 57–66.

Guttman, N. (1953). Operant conditioning, extinction, and periodic reinforcement in relation to concentration of sucrose used as reinforcing agent. *J. exp. Psychol.*, **46**, 213–24.

Hill, W. F., and Spear, N. E. (1962). Resistance to extinction as a joint function of reward magnitude and the spacing of extinction trials. *J. exp. Psychol.*, **64**, 636–9.

Hulse, S. H. (1958). Amount and percentage of reinforcement and duration of goal box confinement in conditioning and extinction. *J. exp. Psychol.*, **56**, 48–57.

Jonckheere, A. R. (1954). A distribution-free *k*-sample test against ordered alternatives. *Biometrika*, **41**, 133–45.

Keller, F. S., and Schoenfield, W. N. (1950). *Principles of Psychology*. New York: Appleton-Century-Crofts.

Kraeling, D. (1961). Analysis of amount of reward as a variable in learning. *J. comp. physiol. Psychol.*, **54**, 560–5.

Lawrence, D. H., and Festinger, L. (1962). *Deterrents and Reinforcement. The Psychology of Insufficient Reward.* Stanford: Stanford University Press.

Logan, F. A. (1960). *Incentive.* Yale University Press.

Siegel, S. (1956). *Nonparametric Statistics.* New York: McGraw-Hill.

Stanley, W. C., and Rowe, M. I. (1954). Extinction by omission of food as a function of goal box confinement. *J. exp. Psychol.*, **48**, 271–4.

Wagner, A. R. (1961). Effects of amount and percentage of reinforcement and number of acquisition trials on conditioning and extinction. *J. exp. Psychol.*, **62**, 234–42.

Weinstock, S. (1958). Acquisition and extinction of a partially reinforced running response at a 24-hour intertrial interval. *J. exp. Psychol.*, **56**, 151–8.

Young, P. T., and Shuford, E. H., Jr. (1955). Quantitative control of motivation through sucrose solutions of different concentrations. *J. comp. physiol. Psychol.*, **48**, 114–8.

Manuscript received 17th January, 1966.

THE EFFECT OF SUPERSEDING SIGNALS*

BY

ROBERT GOTTSDANKER

From the Department of Psychology, University of California, Santa Barbara

Eight adult human subjects were given a step-tracking task in which an occasional second signal within 50, 70, 90, 120, or 240 millisec. called for curtailing or reversing the first command. It was found for inter-signal intervals through 120 millisec. that the shorter the interval the greater was the reduction in amplitude and duration of the majority of responses, with no delay in the effect of the second signal. Where a larger change of response was called for, reversal rather than curtailment, there was a greater effect. A second signal occurring at the 240 millisec. interval (in almost all cases after the start of the response), had no detectable effect. Since the over-all RT was about 180 millisec., it is evident that for at least the first two-thirds of the RT period the initial response is not typically impervious to the effect of a second signal.

Contrary to the expectations of the uncommitted-period version of the hypothesis of substitutive grouping a reversing signal at the 50 millisec. interval did not yield many reversed responses. Moreover this view cannot accommodate the finding that for intervals through 120 millisec., relatively few distributions of response amplitude can be accounted for by the summation of instances of response to the first signal alone and to the second signal alone. It is concluded that for these intervals, there were generally either overlapping responses to the two signals or else unitary responses in which the two signals were grouped to produce a combined effect.

Introduction

A step-tracking task was employed by Vince (1948) in studies which sought to answer the question of whether the "wavy" appearance of records of response in continuous tracking is ascribable to a basic intermittency of function. In this task, the first signal of a pair consisted of a displacement step of the target from the central position, and the second signal the return step after a variable period of time. She found that when the inter-signal interval was 500 millisec. or less, the reaction time (RT) to the second signal was longer than that to the first. The hypothesis was offered, based upon Telford's (1931) concept of a refractory phase of voluntary responses, that the organizing of a response to a signal received while the subject is already busy with a previous signal must wait until the first signal has been processed. In other words, there is an *impervious* period following the first signal. Moreover, a second signal which occurs during the response to the first will also have a somewhat delayed response. However, Davis (1956), in a key-pressing task, found no increase above the normal RT for a second signal which occurred after the end of the RT period for the first.

Vince used the reversal of direction of the initial motion as the indicator of the start of the second response. It may be objected that the response actually must have been under way for some time before the physical requirement for reversal in direction could be effected. In fact, Ellson and Hill (1948), in a virtual repetition of Vince's experiment, obtained similar results which, through detailed analysis of the form of response, they accounted for by the simple *linear addition* in time of the separate responses to the initial displacement and to the return step. Most importantly, their evidence showed that with decrease in the interval between displacement and return signal, there was a reduction in the amplitude and duration of the first response.

* A partial report on the findings was made at the 1963 meeting of the American Psychological Association.

Vince did state ". . . there is some evidence that the first response may be modified slightly by the second stimulus (1948, p. 156)." In a subsequent study (1950), in which curtailing and reversing second signals were used as well as nullifying ones, the estimated incidence of undershot first responses increased (e.g. from 33 to 53 per cent. at the 50 millisec. interval). Further, the usual mode of influence of the second signal in these cases which she infers is much the same as that described by Ellson and Hill, is that the response to the second signal truncates the response which was already started to the first signal. In other words, the responses *overlap*. However, Ellson and Hill contend that linear addition takes place on every trial while in the conclusion to her second study Vince still maintains that overlapped responses are infrequent. Possibly she did not intend to include responses at the shortest intervals, 50 and 100 millisec.

An alternative to overlap as an explanation for reduced-amplitude responses was considered by Vince in the 1950 study: that is that the two signals were grouped and their *effects were combined* from the onset of the response. However, the fact that RT to the first signal had about the same distribution for long and short intervals led against this interpretation. Also relevant was the finding that reduced-amplitude responses went with higher RT to the first signal, possibly indicating that responses which started later were cut off earlier in their course.

Another kind of grouping, in which the response to the first signal is not made, but instead is *substituted for* by the response to the second signal was also considered by Vince. Contrary evidence was found in another part of the 1950 study which showed that subjects were rarely able to start with an initial reversal of the response originally called for, even when the signal for reversal followed the first by as little as 50 millisec. However, Welford (1952) suggests that Vince's finding (in both studies) that the subject made no response at all on about a fifth of the trials at the shortest interval of 50 millisec. points to the operation of substitutive grouping on some trials. He further suggests that reduced-amplitude responses may in some cases be "token" responses which the subject could have prevented had the instructions been completely unambiguous. It might be added that Vince's failure to obtain many initial reversals might have resulted from weak signal-response compatibility; the second signal was not in the reverse direction. A main purpose of the present experiment was to test an explicit version of the hypothesis of substitutive grouping, that of an *uncommitted period* which follows the first signal *on every trial*. With effective instructions not to "acknowledge" signals already past, a sufficiently brief inter-signal interval should produce responses to the second and not to the first signal without exception. Since the period is probably variable from trial to trial, the number of such substitutions should diminish and then vanish as the interval is made longer.

Yet another possibility remains. Responses could overlap, but the effect of the second signal be retarded. That is, psychological refractoriness, even if not valid in terms of temporary imperviousness under these conditions might still be a genuine phenomenon of central interference.

The present study was designed to answer the questions of how generally and to what extent the initial response is affected by the second signal, and what is the nature of the effect when it occurs: overlapping responses (with or without retardation), combinational grouping, or substitutive grouping. The procedure was similar to that of Vince's 1950 study in that the initial signal called for responses of more than one amplitude and direction, but differed in that the second signal called only for curtailment or reversal of the first response, and never for its nullification. Consequently, some recordable response was always required.

It should be noted here that there is much evidence which has been cited as favourable for an impervious period obtained by procedures where different response units such as the two index fingers are used for the first and second responses (e.g. Davis, 1956; Welford, 1959). This procedure was devised to avoid the difficulty of identifying the start of the second response, inherent in the Vince single response-unit procedure. It has been argued that if a delay is found for the response to the second signal even when different hands are used, there must be at least as much central delay with one hand. However, this change in procedure may well introduce new factors and cannot be assumed to represent the sequence of events in the single response-unit case. Only the latter method gives anything like direct evidence on the problem of continuous tracking, which, as Welford points out (1959, p. 207), was "the original stimulus to such studies."

METHOD

Task

Below a horizontal row of five neon lamps there was a pointer having for a handle a $\frac{3}{4}$ in. cube of balsa wood which the subject held between the thumb and forefinger of his right hand. His task was to slide the pointer in a groove to whichever of the lamps was lighted. The middle lamp was considered the "home" position. The near lamps were centred $\frac{1}{4}$ in. from the home centre and the far lamps $\frac{5}{8}$ in. The width of each lamp was $\frac{1}{8}$ in., the pointer had a width of $\frac{1}{32}$ in. A force of about 15 gm. was required to start the pointer in motion.

All trials started with the pointer at the home position. Return movements to that position were of no interest as they could be anticipated, and often took place before the return signal occurred.

Two kinds of trials were employed. On an *uncorrected* trial, with the pointer at the home lamp, one of the other four lamps lighted. Whereupon the subject would slide his pointer as soon as he could to that position. The lamp remained lighted for about 700 millisec. Then the home lamp lighted for about 800 millisec. in preparation for the next trial.

On a *corrected* trial, one of the two far lamps would light for only a brief period, and coincident with its going off, one of two near lamps would light for the remainder of the 700 millisec. before the return to the home position. In effect, the subject was given a new command which either curtailed or reversed the initial command. The brief intervals employed on the corrected trials were: 50, 70, 90, 120, and 240 millisec. Needless to say, the subject was not informed in advance as to whether any particular trial would be corrected or the response which would be required.

Instructions

Following are the significant portions of the instructions, which were read aloud to the subject. Those shown in italics were not only read before the start of the experiment as a whole but also before every test session:

"The target moves in jumps from place to place. Your task is to stay on target as much as you can. *You should try to be where the target is and not be concerned with where it was or where it will be.* You must move the control as quickly as you can. It isn't necessary for the pointer to be precisely in the middle of the target mark; just keep the pointer within the lighted band."

As is evident, there was great emphasis placed upon contemporeneity of response with signal and on the avoidance of anticipating.

Programme

Signals were sequenced by the combined action of two stepper switches. However, the short intervals on the corrected trials were precisely determined by a time-interval generator triggered by the mechanism for the selection of signals. There were seven times as many uncorrected as corrected trials. On the uncorrected trials there were very close to the same number of trials for each of the four positions; on the corrected trials there were the same number for each type of correction (e.g. initial-left, curtail) and for each of the inter-signals intervals.

The programmed sequence consisted of 160 trials with the different possible events in scrambled order. On each of the four test days, the programme was gone through five times, giving each subject a grand total of 2,800 uncorrected trials and 400 corrected trials. During a test day the trials were given in nine blocks, each starting at a different point in the programme (to prevent memorizing). There was a 2-min. rest interval between successive blocks. The daily session lasted about an hour. A practice session of the same length was given the day before the tests started.

Measurements

Signals and responses were recorded continuously on a two-channel Brush galvanometric recorder. The output of the signalling apparatus actuated one of the pens. The subject's control handle was attached to a linear potentiometer, the continuous output of which actuated the second pen. The recorder paper moved at a speed of 125 mm./sec., allowing determination of points in time to the nearest 10 millisec. Amplitude was recorded on almost the same scale as the response itself. The ⅝ in. (15·9 mm.) between the centre lamp and either far lamp showing as 15·3 mm. on the recording paper. Measurement was to the nearest millimeter on the record.

Three basic measures were obtained: reaction time, amplitude of initial response, and duration of initial response. The method of measurement is shown in Figure 1. Independent measurements of all of the records were made by two readers who were unaware of the significance in the experiment of any of the measures, and differences were reconciled by one of the readers. Also, it should be noted that the trials for each subject were scored in sequential order, not segregated as to experimental condition. On the signals to the right, an apparent initial pulse to the left is seen. This is simply a recording artefact. The lamp was lighted at the bottom point of the stroke. Distances of signal deflection on the chart are not proportional to display distances.

Reaction time (**RT** in Fig. 1) is the interval between the onset of the signal (first signal in the case of a corrected trial) and the onset of the response. It will be noted that the signals in Figure 1 are: (A) to the far left and uncorrected; (B) to the near right and uncorrected; (C) first signal to the far left, reversed after 50 millisec; (D) first signal to the far right, curtailed after 90 millisec.

Amplitude (**A** in Fig. 1) is the height of the trace of response (or initial response for corrected trials). In the results it is presented in terms of the percentage of the requirement of the far response, i.e. of the ⅝ in. between the home lamp and a far lamp.

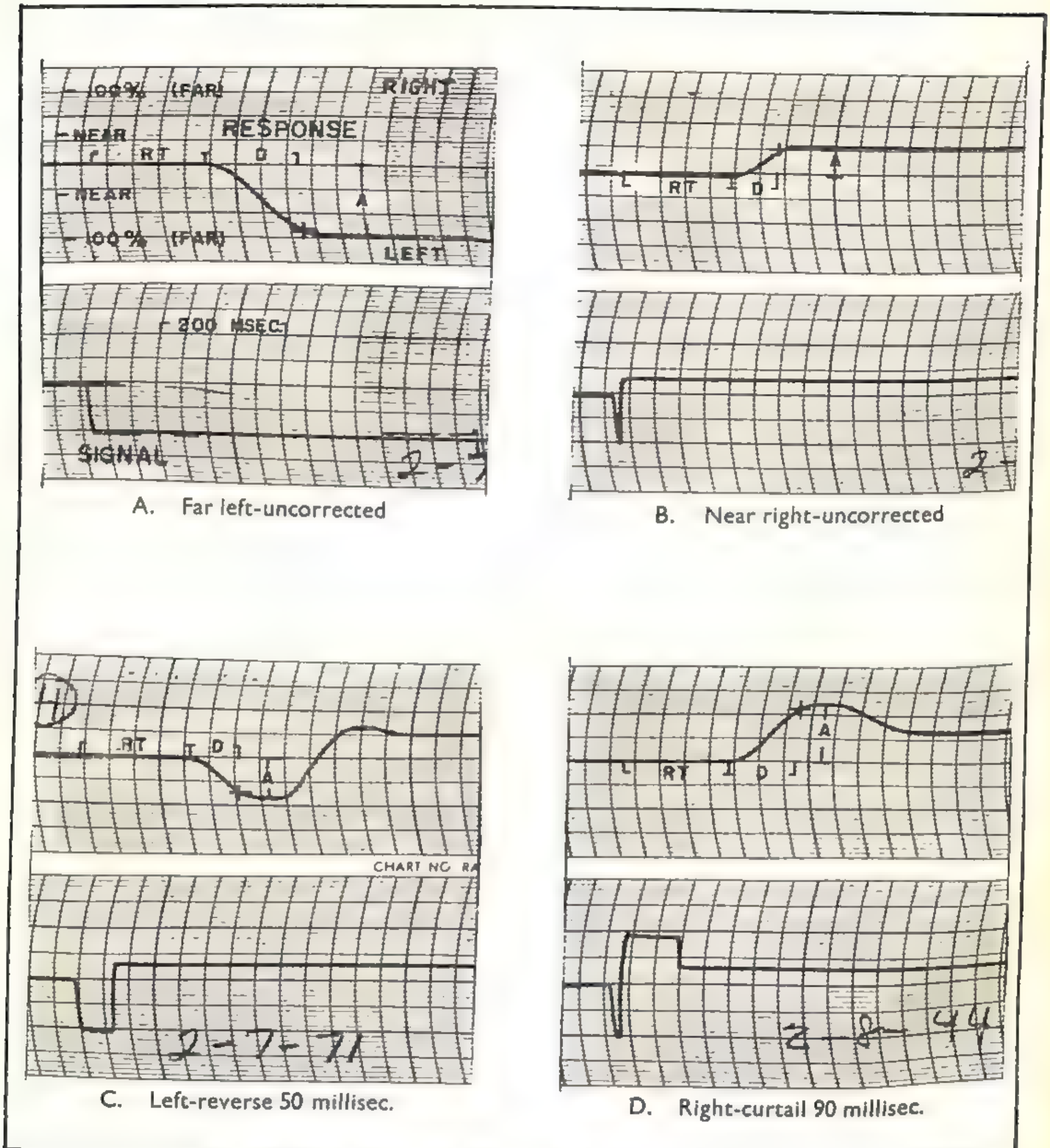
Because of the asymptotic nature of many of the responses (note Fig. 1D) it was not found possible to get reliable agreement between scorers as to the point in time when the maximum amplitude was reached. Consequently the measure of *Duration* employed here was the time interval between the onset of movement and the instant the trace reached within 1 mm. of the maximum amplitude. Examples are shown in Figure 1, labelled "**D**."

Irregular types of response obtained are shown in Figure 2. First, in 2A, there is seen a "wrong-direction" response. Instances in which the initial response was in the direction opposite that prescribed by the signal—or the first signal—were so classified. In 2B there is seen another wrong-direction movement, but which was immediately corrected. In 2C there is seen an example of an amorphous response. It may be noted that while all of the other examples of responses in Figures 1 and 2 have a single inflection, at the instant of maximum velocity, here there are at least three inflections. The response in Figure 2A has an exceptionally short RT, 70 millisec. Extremely long RTs were occasionally obtained, up to 500 millisec. The example in 2D shows a RT of 330 millisec.

Almost all of the corrected trials were scored. The exceptions were those which occurred at the outset of a block of trials. Because of their large number a random sample was taken of the uncorrected trials. A few other trials, both corrected and uncorrected, were lost because of such mishaps as clogged pens. For each of the four uncorrected situations, there was an average of about 60 trials scored for each subject. For each of the corrected situations (e.g. initial signal to right, followed at 70 millisec. by reversal signal to left) there were ordinarily about 20 trials for each subject, in almost all cases between 15 and 25.

Subjects

Eight men participated in the experiment. Four were members of the project, and ranged in age from 21 to 45. The other four were paid 16-year-old high-school students.



A. Far left-uncorrected

B. Near right-uncorrected

C. Left-reverse 50 millisec.

D. Right-curtail 90 millisec.

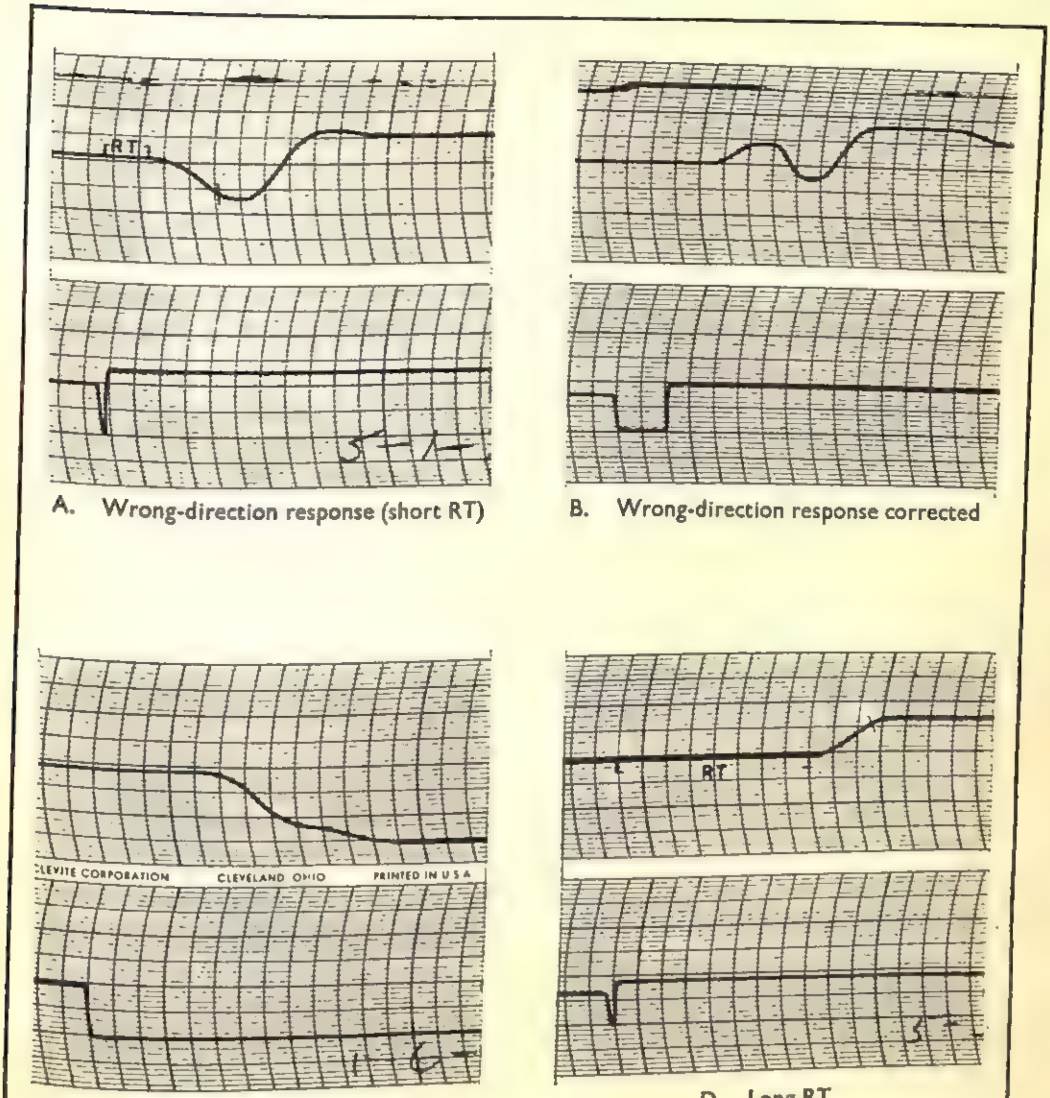
FIGURE 1

Signals for various conditions, typical responses, and measurements. (The large hand-written numbers are trial identifications.)

Results

Reaction Time

Figure 3 shows the mean values for the eight subjects of the median RTs for the various conditions. Mean values of Q (semi-interquartile range) are also presented. All of the scored trials were used for this analysis. It is first of all seen that median RT is quite short for this kind of task. Values centre around 180 millisec. This is considerably shorter than the RTs found by Vince and by Ellson and Hill. Waiting



A. Wrong-direction response (short RT)

B. Wrong-direction response corrected

C. Amorphous response

D. Long RT

FIGURE 2

Examples of irregular types of response.

for the second signal doesn't seem to have occurred, a possible explanation of grouping mentioned by Welford (1952).

No consistent differences are related to length of interval, so if the second signal did influence the initial organization of response, it did so without increasing RT. No consistent differences in Q emerge either. A consistently shorter RT to the right is noted. It would seem probable that this indicates no more than the usual preference for right-going motion.

It may be added that at the 240 millisec. interval, the response had been started on 94 per cent. of the trials before the appearance of the second signal.

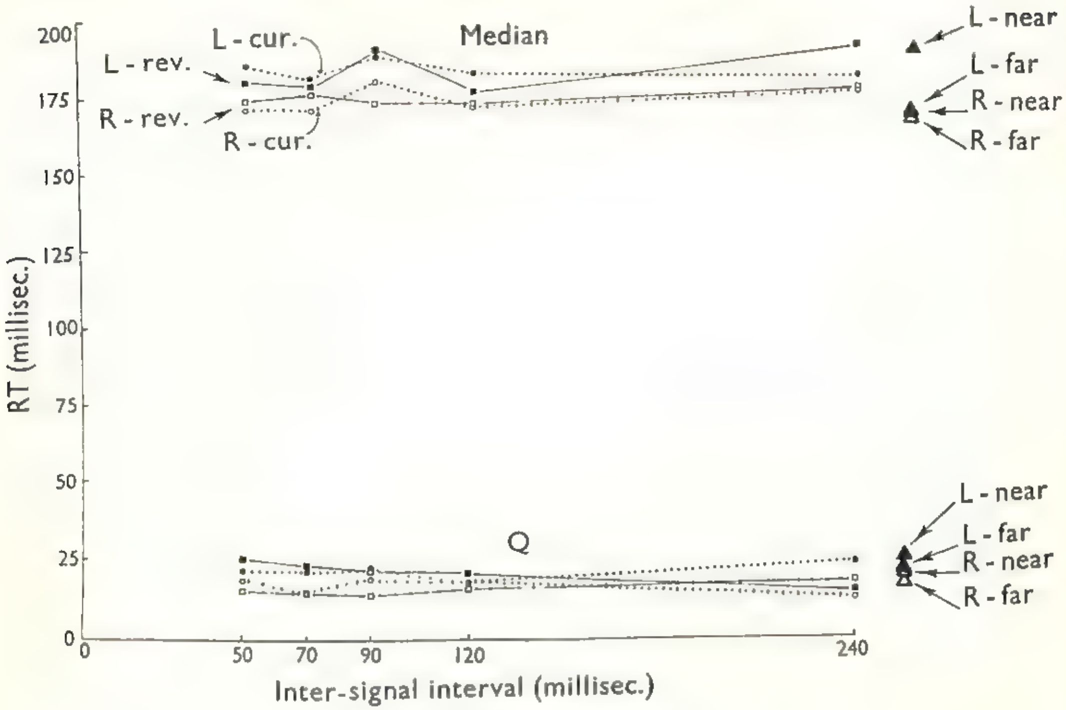


FIGURE 3

Mean values of median RTs and of Q (semi-interquartile range).

Wrong-direction responses

These were right-direction responses in respect to the second signal if a reversal was called for. However, the reference taken was the first signal of the pair or the single uncorrected signal. It was found first that there was a great deal of variability among subjects in the number of such responses. Whereas subject 2 and subject 3 made fewer than one per cent. wrong-direction responses, subject 8 made over 13 per cent.

Table I shows the average percentage for the eight subjects of wrong-direction responses for the various conditions. It is seen that they were never in the majority, even for the conditions which might be expected to favour them. Consequently, it must be concluded that if the second signal affects the response from the outset, its contribution is less than that of the first signal. Vince (1950) obtained substantially the same finding.

Now attention may be given to whether these responses were most prevalent for the reversal, short-interval condition. Comparisons with the uncorrected trials, where only a far signal was given, are most to the point. First, it should be noted that on the uncorrected trials there were more reversals on signals to the left: another example of the right-going tendency. Next, it is evident that for the left-reversal condition, 50 millisec. interval, there were more wrong-direction responses than for the uncorrected condition (15·3 per cent. vs. 4·2 per cent.). Within the much smaller number of wrong-direction responses for the trials with the first signal to the right, the same difference is seen (5·6 per cent. vs. 0·6 per cent.).

These over-all results are reflected in the data of the individual subjects. In all, 16 comparisons may be made between the reversal condition with the 50 millisec.

interval and the far uncorrected condition in terms of percentage of wrong-direction responses. In 14 of these comparisons, there was a higher number for the reversal condition. By way of contrast, in the comparison of the curtail condition (again at the 50 millisec. interval) with the far uncorrected condition, there were slightly more cases of a larger number of reversals for the uncorrected condition.

TABLE I

MEAN PERCENTAGE OF WRONG-DIRECTION RESPONSES ACCORDING TO EXPERIMENTAL CONDITION

Corrected	*Interval (millisec.)* 50	70	90	120	240	*Uncorrected*	
Left-Reverse	15·3	12·3	5·9	8·8	2·2	Left-far 4·2	Left-near 6·6
Left-Curtail ..	3·6	5·2	6·1	4·6	4·0		
Right-Reverse	5·6	1·0	1·6	0·5	1·6	Right-far 0·6	Right-near 2·0
Right-Curtail	2·8	1·6	3·8	0·8	0·8		

However, it should not be surmised that the wrong-direction responses in the reversal condition were controlled by the second signal alone. It will be shown in the next section that they tended to be of lesser amplitude than responses made to the second signal when it was presented uncorrected.

Amorphous responses

Great variability among subjects was found in their percentage of amorphous responses, those in which there was more than one inflection in the response curve. The range was from 1·8 to 30·6 per cent.

The only unequivocal difference was that between the far and near uncorrected patterns. There were in excess of six times as many amorphous responses on the far patterns. This is as would be expected, the chance for inflection being much greater for a higher-amplitude response. None of the eight subjects showed a greater number of amorphous responses for the near signal, on either side.

Differences found for the corrected patterns were inconsistent throughout. Overall about the same percentage of amorphous responses was found for each of the four shortest intervals.

Amplitude

The most important data of the present study relating to the extent of influence of the second signal are the values of amplitude of response. Mean values of median amplitudes for the eight subjects are shown in Figure 4. Also shown are the mean values of the median durations. The lower curves show mean values of Q. It should first be noted that the points in Figure 4 included only those responses in the direction of the first or uncorrected signal and those on which there was no possibility that a "safety-first" technique was adopted, in which a short response was made, which was extended if no correction signal was immediately forthcoming: i.e. wrong-direction and amorphous responses were excluded. For each of the four uncorrected conditions between 27 and 97 trials entered the analysis for each subject. For the corrected situations, there was but little reduction in the number of trials for some subjects but more drastic reductions (down to 12 trials or even fewer) occurred for

other subjects. Excluded responses have already been considered in previous sections and the effect of their elimination will be discussed presently.

Required amplitudes were 100 per cent. for the uncorrected far signals and 40 per cent. for the near signals. It is seen that the far signals tended to be slightly undershot, with an over-all mean amplitude of about 95 per cent., and the near signals slightly overshot, with an over-all amplitude of about 45 per cent. It is evident that as interval increases, amplitude does likewise without exception. The longest interval, 240 millisec., gives medians almost identical with those for the uncorrected condition.

This increase of amplitude between successive intervals is borne out in the data of the individual subjects. Since for each correction-condition there are four comparisons between successive intervals, for the four conditions there are 16 such comparisons. Over the eight subjects, then, there are 128 comparisons between successive intervals. In only 17 instances did the median amplitude of response fail to rise from one interval to the next longer. Seven of these failures occurred for a single subject.

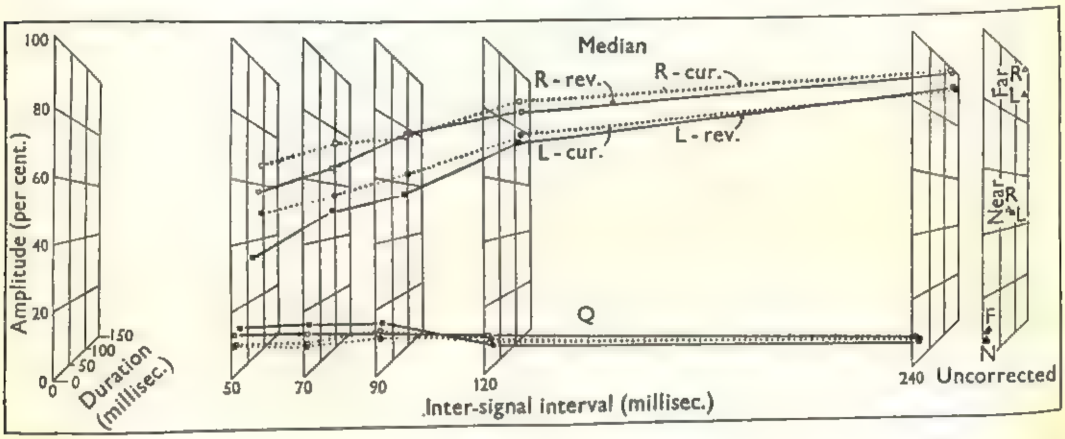


FIGURE 4

Mean values of median amplitudes and durations, and corresponding mean values of Q.

Two other kinds of differences are notable. Higher amplitude responses were consistently made to the right, although this is not so clearly the case for responses to the near signal. Again, the most likely explanation is the preference for left-to-right motion. More important for the present consideration are the lower amplitudes for the reversal condition than for the curtailment condition at the two shortest intervals. Separate comparisons were made for initial signal to left and to right on these two intervals for the eight subjects: 32 comparisons in all. Of these comparisons, 25 show lower amplitude for the reversal condition.

There is some indication that variability, as measured by Q, is elevated for the shorter intervals, and especially for the two shortest intervals under the reversal condition. Any precise comparison of variabilities becomes questionable because of the small number of cases in each distribution.

It is obvious that if the wrong-direction responses had been included in the determination of median amplitudes, the effect would have been to *enhance* the relation between interval and amplitude for the reversal condition, since a disproportionate number of these responses occurred at the shortest intervals. The necessary elimination of amorphous responses could not have introduced any artefact.

There was no disproportionate number of responses eliminated for any correction condition.

Perhaps the case has not been made convincingly that individual amplitudes for intervals shorter than 240 millisec. were not generally determined by the first signal alone. Might it not be possible that the initial response was only "modified slightly" by the second signal, either in terms of the extent of effect or the percentage of cases in which there was any effect? This does not seem conceivable in light of the degree to which median amplitudes for the different intervals deviated from the 95 per cent. value for the uncorrected far responses: 50 per cent. for the 50 millisec. interval, 60 per cent. for the 70 millisec. interval, and 70 per cent. for the 90 millisec. interval.

Comparison of individual distributions (one subject, one correction condition, one interval) with the median amplitude for the far signal uncorrected only reinforces the foregoing conclusion. If a subject responds exactly the same on each trial to a given condition-interval combination as he does to the first signal when uncorrected, 50 per cent. of his responses will have amplitudes *greater* than the median amplitude for the uncorrected condition. By extension it may be argued that if 25 per cent. of the amplitudes are above the median, he will have made 50 per cent. of his responses just as he did to the uncorrected first signal. Actually, this is the highest possible estimate. Many or even all of the responses could be of lower amplitude than they would have been without the second signal. Nevertheless, this criterion was selected in order to detect all distributions in which there was a *possibility* that half or more of the responses were made to the first signal alone. Wrong-direction responses were included in this analysis, making the distributions slightly larger in some cases than those used for finding median amplitudes. There were a few responses to the near signal alone which were greater in amplitude than the median for the far uncorrected condition (fewer than 0·5 per cent.). The effect of such responses on the classification of corrected trials would be inconsequential, and could only raise the estimate of responses possibly attributable to the first signal alone.

Results of these comparisons are shown in Table II, under the heading *S1*. It is seen that for the 240 millisec. interval, 30 of the 32 distributions met the criterion. It is likely that the two which did not were the result of sampling error as it was shown previously that the over-all median amplitude was not lower than for the uncorrected signal. For the 120 millisec. interval, 17 of the 32 distributions meet the criterion, and for the three shortest intervals, only seven of the 96 distributions. Thus, for intervals through 90 millisec. there was only a small proportion of distributions on which half or more of the responses could possibly be attributable to the first signal alone.

It is held by the uncommitted-period version of the hypothesis of substitutive grouping that those responses not made to the first signal were made to the second, rather than to a combination of the two. Actually, two questions are involved: (1) Are there distributions in which all of the amplitudes may be accounted for by responses to one or the other signal? (2) If so, do these distributions contain responses attributable to each of the signals?

If 50 per cent. of the amplitudes are beyond either the median amplitude for the first signal when uncorrected or beyond the median amplitude of the second signal when uncorrected, all of the responses may be attributable to responses made to one or both signals, without combinational effects. Since there doubtless were sampling errors, this is too strict a criterion to apply. Instead, the criterion selected was that of 40 per cent. of the amplitudes falling beyond either median. For the first signal *beyond*, as before means of higher amplitude. For a curtailing second

signal, it means of lower amplitude; for a reversing second signal, it means of higher amplitude in the reverse direction.

Results of this analysis are shown in Table II under the heading *Ei*. At the 240 millisec. interval the majority of distributions (27 out of 32) meet this criterion. At the four other intervals it is met by only 19 of the 128 distributions. Of the 46 instances in which the criterion is met, there are 45 where it is satisfied by amplitudes beyond one of the medians alone. It is even typical for there to be *no* trials beyond one of the two medians (39 of the 46 instances). For the separate distributions there was not a division of responses to each of the two signals.

TABLE II

Number of Distributions with 25 per cent. or More Amplitudes Beyond Median for First Signal Uncorrected (S1), Beyond Median for Second Signal Uncorrected (S2), or with 40 per cent. or More Beyond Either Median (Ei)

Correction Condition	*Interval (millisec.)*														
	50			*70*			*90*			*120*			*240*		
	S1	*Ei*	*S2*	*S1*	*Ei*	*S2*	*S1*	*Ei*	*S2*	*S1*	*Ei*	*S2*	*S1*	*Ei*	*S2*
Left-Reverse ..	0	0	0	0	0	0	2	1	0	4	2	0	8	8	0
Right-reverse ..	1	1	0	1	1	0	1	1	0	6	1	0	8	6	0
Left-Curtail ..	0	4	6	0	2	4	0	2	2	3	2	0	7	6	0
Right-Curtail ..	1	1	2	0	0	0	1	0	2	4	1	0	7	7	0
Total of the 32 possible ..	2	6	8	1	3	4	4	4	4	17	6	0	30	27	0

Now it is appropriate to ask the same question concerning the control by the second signal as was asked relative to the first. In how many distributions may half or more of the responses possibly be attributed to the second signal alone. The 25 per cent. criterion was applied to amplitudes beyond the median for the second signal. Results of these comparisons are shown in Table II under the heading *S2*. There was a small number of responses to the far signal alone lower in amplitude than the median for the near signal alone (less than 3 per cent.). The effect of such responses on the classification of corrected trials would be inconsequential, and could only raise the estimate of responses possibly attributable to the second signal alone.

There are no distributions which meet the criterion for either reversal condition. Even when just the wrong direction responses were considered at the 50 and 70 millisec. intervals, only 11 of the 48 were beyond the median for the second signal. The instances of reversal responses described previously do not in the majority of cases represent sole control by the second signal. For the two curtail conditions, no distribution meets this criterion for either the 120 or 240 millisec. interval. For the three shortest intervals, there are a substantial number of distributions which qualify, especially for signals to the left at the 50 millisec. interval.

The evidence in Table II, then, is that except at the 120 and 240 millisec. intervals there are relatively few distributions of amplitude where even half of the responses may possibly be attributed solely to the first signal. This extent of control by the second signal alone is seen only for the curtail condition at the shortest intervals. The situation was not often found in which the bulk of amplitudes could be accounted for by response to one signal or the other. The majority of responses at intervals through 90 millisec. and almost half at 120 millisec. must represent the combined effect of the two signals.

Duration

Referring back to Figure 4, it may be seen that duration of response increases with amplitude, generally. This is in accord with observations of previous investigators. Increase in duration with interval is almost as impressive as that of amplitude. Only one exception occurs in the over-all records: the mean duration on the right-curtail dipping from 105 millisec. at the 70 millisec. interval to 104 millisec. at the 90 millisec. interval. Of the 128 comparisons on the curves of individual subjects, there were only 18 failures for duration to rise between successive intervals.

It is noted that the very longest median durations are less than 150 millisec., a bit shorter than the over-all median RT. Thus, there seems no possibility that the subjects used the technique of breaking responses into an initial and a corrected portion. This may conceivably be the case for some of the amorphous responses. However, it will be recalled that these responses were not used in the analyses of median amplitude or duration.

It has already been demonstrated in both the amplitude and duration data that the subject was not impervious to the second signal. Still, there may have been an interference effect of the first signal. This would make the time between the second signal and its first noticeable effect (i.e. its RT) longer than a normal RT. It was not possible to determine the *initial* effect of the second signal in the present study. However, this effect must have occurred *before* the point in time used for determination of duration, the instant at which the response reached within 1 mm. of its maximum amplitude. Hence, if this point is reached within an amount of time not markedly longer than the normal RT, it cannot be held that there was a delay in the influence of the second signal. This holds true regardless of the kind of correcting signal. Since the duration values for the left-reverse condition are generally lower than those for the three other correction conditions, they have been selected for examination.

The mean value of median duration for the 50 millisec. interval was 70 millisec. When coupled with the 180 millisec. RT for this interval, it is indicated that the criterion amplitude (1 mm. from peak) was reached within 200 millisec. after the second signal (180 − 50 + 70). Similar comparisons for the next three intervals show: 70 millisec. interval, RT of 179 millisec. and time to criterion amplitude, 204 millisec.; 90 millisec. interval, RT of 190 millisec. and time to criterion amplitude, 195 millisec.; 120 millisec. interval, RT of 177 millisec. and time to criterion amplitude, 177 millisec. Thus, it cannot be said that there was any real increase over a normal RT in the length of time required for the second signal to have an effect.

Correlation between amplitude and reaction time

Vince (1950) found that low-amplitude responses at the short intervals were most frequent when RT to the first signal was high. In order to determine whether there was the same tendency in the present experiment, rank-order correlation coefficients (rho) were obtained between amplitude and RT. As can be seen in Table III, which presents the median values for the eight subjects, the negative relation generally did hold for intervals through 120 millisec., but not for the 240 millisec. interval nor for the uncorrected conditions.

Discussion

What is most evident is that there was generally a graded effect of the second signal. For inter-signal intervals of 50 through 120 millisec. there was greater reduction in amplitude and duration of the initial response as the inter-signal interval

was briefer. For the very short intervals, a reversing second signal brought about a greater reduction than did a curtailing signal. There was no effect of a second signal which occurred at the 240 millisec. interval, by which time the response almost always had been started.

It is hardly necessary to point out that these findings make untenable the existence of an invariable condition of imperviousness during the RT period. It is important to note that an effect of the second signal for intervals through 90 millisec. was not a rarity; it was the usual occurrence. On very few distributions of the individual subjects could even half of the responses be ascribed, by the most generous standards, to the first signal alone.

TABLE III

Median Values for Rank Order Correlation (Rho) Between RT and Amplitude

Corrected	*Interval (millisec.)*					*Uncorrected*	
	50	*70*	*90*	*120*	*240*		
Left-Reverse	−0·43	−0·32	−0·32	−0·32	−0·04	Left-far −0·01	Left-near −0·08
Left-Curtail . .	−0·18	−0·28	−0·19	−0·26	0·14		
Right-Reverse	−0·48	−0·34	−0·37	−0·24	−0·07	Right-far 0·01	Right-near −0·02
Right-Curtail	−0·26	−0·35	−0·23	−0·08	0·15		

Whether there were any responses at the short intervals which were uninfluenced by the second signal cannot be decided with certainty. There is no way of knowing what the amplitude of an individual response would have been had there been no second signal. However, this is not an issue in the present inquiry. It cannot be doubted that a subject is capable of disregarding the second signal; the real question is whether he is capable of utilizing it. There is no room for doubt on this score.

As for the possibility that the first signal had at least the effect of increasing RT to the second signal the evidence—although admittedly indirect—is negative. The influence of the second signal on the initial response did not appear to be delayed beyond a normal RT.

The uncommitted-period version of the hypothesis of substitutive grouping leads to the expectation that with extremely brief inter-signal intervals, the response will be controlled entirely by the second signal. There was some evidence this could happen under the curtailing condition; but not under the reversing condition. Of course, an even briefer inter-signal interval might produce a majority of responses which are accountable for by the second signal. However, the question would then have to be considered of whether there was an actual perception of succession between signals. It seems implicit in the statements of the substitutive grouping hypothesis that while there is awareness of succession, the response is nevertheless made to the second signal. The aspect of this hypothesis which was not at all realized was the determination of response by one signal or the other. No large number of distributions could be attributed to summation of instances of responses to the first signal alone and to the second signal alone.

A firm demonstration that there indeed were responses which were determined solely by the second signal could not be accommodated by the hypothesis of overlapping responses to the two signals. Unfortunately, the present evidence is inconclusive on this question.

The hypothesis of overlapping responses thus remains as an explanation for which there was a good deal of positive evidence and no strong contradictory evidence. A competing hypothesis which remains tenable is that of combinational grouping. It would thus be maintained that for some period following the first signal, a second signal, if introduced, will influence the initial response from the outset, in some instances possibly to the exclusion of the influence of the first signal. There still might exist a state of imperviousness to a second signal in the final portion of the RT period. Although, the negative correlation between amplitude and RT, without increase in the variability of RT at short intervals, gives more support for overlapping responses, as Vince (1950) points out this is not unequivocal evidence. A test between the two hypotheses would be to determine whether the effect of the second signal is found at the very start of the response, or as held by the hypothesis of overlapping responses, only after a period of one RT beyond the second signal. It is planned in a future study to make a continuous record of instantaneous acceleration during responses to corrected and uncorrected signals to obtain evidence on this question. Preliminary work has already been undertaken (Gottsdanker, 1963).

Especial thanks are given to Dr. L. Braley and Mr. J. Senders, who assisted in the planning and conduct of the experiment. Appreciation is also due to the many psychologists with whom the writer discussed preliminary analyses he had made of the data. The project was supported by the United States Air Force Office of Scientific Research, Contract AF 49 (638)—730, with additional help from research funds granted by the University of California.

REFERENCES

DAVIS, R. (1956). The limits of the "psychological refractory period." *Quart. J. exp. Psychol.*, **8**, 24–38.

ELLSON, D., and HILL, H. (1948). The interaction of responses to step function stimuli. I. Opposed steps of constant amplitude. *Air Materiel Command. Memorandum Report MCREXD—694–2P*. Pp. 37.

GOTTSDANKER, R. M. (1963). How responses are influenced by superseding signals. *Paper read to Psychonomic Society.*

TELFORD, C. (1931). The refractory phase of voluntary and associative responses. *J. exp. Psychol.*, **14**, 1–36.

VINCE, M. A. (1948). The intermittency of control movements and the psychological refractory period. *Brit. J. Psychol.*, **38**, 149–57.

VINCE, M. A. (1950). Some exceptions to the psychological refractory period in unskilled manual responses. *Med. Res. Council, A.P.U. Report No. 124/50.*

WELFORD, A. T. (1952). The "psychological refractory period" and the timing of high-speed performance—a review and a theory. *Brit. J. Psychol.*, **43**, 2–19.

WELFORD, A. T. (1959). Evidence of a single-channel decision mechanism limiting performance in a serial reaction task. *Quart. J. exp. Psychol.*, **11**, 193–210.

Manuscript received 6th February, 1966.

SHORTER ARTICLES AND NOTES

DISRUPTION OF A POSITION-REVERSAL LEARNING SET IN THE RAT

BY

S. D. DALRYMPLE and R. STRETCH

From the Department of Psychology, University of Alberta, Edmonton

Nineteen rats were required to learn a series of 15 position-reversal problems in a T-maze (Phase A) and were then assigned at random to two groups. The acquired position-reversal set (PRLS) was then subjected to disruption over 80 trials in which different reward treatments operated for each group (Phase B). For a further 150 trials following the disruptive phase, the initial conditions were reinstated (Phase C). Results showed that rats receiving non-reinforcement during Phase B returned to the asymptotic performance recorded in Phase A within 2–3 reversal problems while those receiving 20 per cent. random reinforcement during Phase B failed to re-acquire the PRLS within 15 reversal problems. The results indicate that random 20 per cent. reinforcement, applied after establishment of a PRLS in rats, displaces by persistent random-responding, the "Win-stay-Lose-shift" hypothesis appropriate for PRLS attainment.

Introduction

In serial position-reversal experiments, subjects learn a number of positional discriminations in which the location of reward is changed from problem to problem in an alternating sequence. Although several investigators (North, 1950*a*; 1950*b*; Dufort, Guttman and Kimble, 1954; Pubols, 1957; Stretch, McGonigle and Rodger, 1963; Stretch, McGonigle and Morton, 1964), have reported a progressive improvement in reversal performance over a series of such problems, a search of the literature reveals that there have been no experiments carried out to examine the effects of procedures designed to disrupt performance of a position-reversal learning set (PRLS). The present study represents an approach to this problem. After establishment of a PRLS, the rats were assigned to two treatments: (*i*) 20 per cent. intermittent reinforcement and (*ii*) omission of reinforcement. These conditions were in effect for 80 trials and the original reward contingency was then reinstated. Performance might then be affected adversely in one of three ways: (*a*) the PRLS could be disrupted and the instrumental (running) response maintained; (*b*) the running response could be disrupted and the set retained; and (*c*) both might be impaired.

Method

Apparatus

The apparatus, which has been described in detail in a previous report (Stretch, McGonigle and Rodger, 1963), was an enclosed T-maze consisting of a start box, runway and choice-point, two arms, and two goal boxes. Doors were made of brown plywood and the rest of the maze was grey. Each goal-box entrance was covered by a door that swung inwards only, thus preventing retracing; guillotine doors separated the arms from the runway. White noise was used to mask incidental sounds and illumination was provided by a 100-w. bulb suspended above the choice-point.

Subjects

The subjects were 19 male albino rats (mean age: 75 days); they had not been used previously for experiment. All of them were habituated to handling during the establishment of a 23-hr. food deprivation schedule.

Experimental design

Subjects were required to learn a series of position reversals in the T-maze. Each received 10 trials a day throughout the experiment, the correct position being reversed each day. The experiment was divided into three procedural phases:

Acquisition (A), "Disruption" (B), and Re-acquisition (C) of the PRLS. During the acquisition phase each animal received 15 position-reversal problems. For the "disruption" phase, they were assigned at random to two groups: Group RR (N = 10), rewarded at random on two of the 10 daily trials and Group NR (N = 9) receiving no reward throughout this period. During this phase, 80 trials (10-per-day) were given and for scoring purposes, two criteria were employed: (*i*) the "correct" position alternated from day to day so that "errors" could be compared with the positional responses during acquisition, and (*ii*) if any rat failed to enter one of the goal-arms within 30 sec., it was removed and a "time-error" was noted. During Re-acquisition (C), each rat received 150 trials under conditions identical to those of Acquisition, i.e. it was possible for it to obtain reinforcement on every trial providing the correct positional response, which alternated for each block of 10 trials, was made.

Procedure

A 23-hr. food deprivation schedule was established during a period of 10 days prior to maze-training; water was available in the home cages at all times. Ten pre-training trials were given over 4 days and, on the basis of the last five of these, the position preference of each animal was determined. The following day, discrimination training began according to the above design; on problem 1, all subjects were required to run against their positional preferences, i.e. reward (wet mash) was located in the non-preferred goal box. The procedure on a given trial was as follows: the experimenter placed the rat in the start box and raised the door, lowering it again when the rat entered the runway. After it had entered one of the goal-arms, the two guillotine doors at the choice-point were lowered simultaneously. A non-correction procedure was used. If the rat made a correct response, it was required to push through the goal-box door and was permitted to eat for 10 sec. If it entered the negative goal-arm, it was detained for 10 sec., the goal-box door being locked on that side. On completion of a trial the rat was removed to await its next trial; the inter-trial interval throughout the experiment was about 2 min. On each trial the experimenter recorded whether the subject was correct or not and, during the "disruption" phase, noted any time-errors that occurred.

Results

All subjects received the same treatment during the acquisition phase and thereafter were assigned at random to two treatments for the subsequent (disruptive) phase. For purposes of analysing the data, however, the groups were treated at acquisition as if they were already assigned to their respective conditions. The mean percentage of correct responses per trial for blocks of five problems during acquisition (A), a block of eight "problems" during disruption (B), and for blocks of five problems during re-acquisition (C) are plotted in Figure 1 which serves to summarize the results as a whole. During acquisition (A) all subjects acquired with roughly equal facility position-reversal learning sets; no significant difference in terms of the rate of acquisition of the PRLS was found between animals that were then subjected to separate treatments during Phase B.

During the "disruptive" phase (B), both groups tended to respond essentially at random (see Fig. 1), making approximately 50 per cent. of their responses to each goal-arm. But, whereas the group that received random 20 per cent. reinforcement continued in this way throughout the "disruptive" phase, i.e. 20 per cent. intermittent reinforcement was evidently adequate to maintain the instrumental running response, the group subjected to non-reinforcement tended to accumulate "time-errors," i.e. they failed to enter a goal-arm within 30 sec.; as a result, the mean percentages of "correct" response is on average lower (by 10–15 per cent.) for this group during Phase B.

In order to determine the difference in performance between acquisition (A) and re-acquisition (C) for the two groups, the Wilcoxon Matched-Pairs Signed Ranks test (Siegel, 1956) was applied to the last five acquisition problems and the first five re-acquisition problems for each group independently. For those subjects rewarded randomly during the disruptive phase (Group RR), a deterioration in performance was found ($T = 0$; $p < 0{\cdot}01$). However, the group receiving non-reinforcement (Group NR) showed improvement in performance after the reinstatement of the original reward contingency ($T = 0$; $p < 0{\cdot}01$).

The differences in performance between the two groups during re-acquisition are clearly evident from Figure 1 (Phase C), and were tested by means of the Mann-Whitney

test. The result indicated that Group RR was significantly inferior to Group NR throughout this phase ($U = 0$; n_1, 10; n_2 9; $p < 0{\cdot}001$) although there is evidence of improvement in the performance of Group RR as Phase C continued. Significant differences in performance remained, however, throughout the 15 problems comprising re-acquisition.

Discussion

The main result to emerge from this study is the fact that those rats receiving non-reinforcement during the second phase of the experiment returned within the first 20 trials of re-acquisition to the level of performance they had attained at the end of the acquisition phase. On the other hand, the intermittently-rewarded group was significantly less efficient in re-acquiring the PRLS and, after 15 re-acquisition problems, had failed to attain the level of performance formerly reached in acquisition.

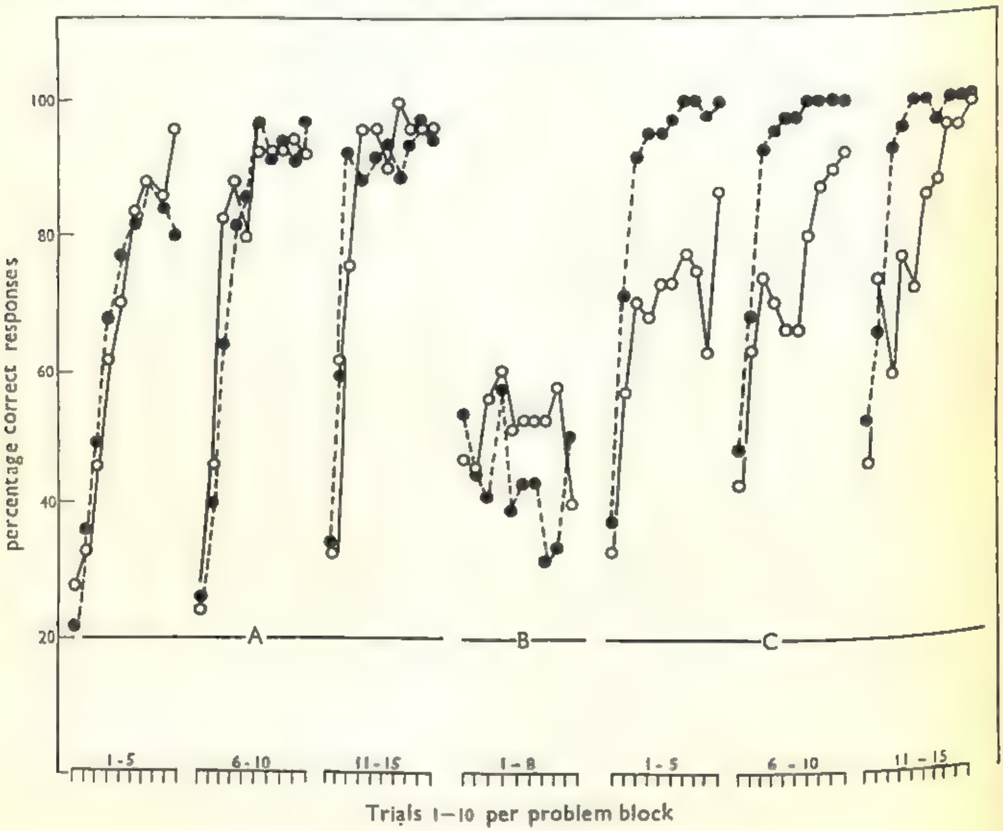


FIGURE 1

Mean percentages of correct responses per trial for blocks of five problems during Acquisition (A), one block of eight "problems" during "Disruption" (B), and for blocks of five problems during Re-acquisition (C). The broken line denotes Group NR and the unbroken line Group RR.

It was observed during the course of the "disruptive" phase, that the running times of the non-rewarded group gradually increased. By the fifth "problem" of this phase, time-errors had begun to occur and it was evident that the instrumental response of running for rats in this condition had been impaired. For the randomly-reinforced 20 per cent. group, however, the reduced frequency of reward was adequate to maintain the running response. Of the three possible forms of disruption described in the introduction, it would appear that Group NR suffered disruption of the running response but retained the discrimination reversal set without loss of efficiency. The running response was maintained in the case of Group RR but the reversal set evidently underwent

severe disruption and was not re-established at its former level of efficiency. More specifically, it appears as if the "Win-stay-Lose-shift" hypothesis (Levine, 1959; Pubols, 1962), appropriate for efficient PRLS performance, was displaced by an essentially "random-responding" hypothesis as a result of the intermittent reward condition during phase B and that the latter continued to influence performance even though the original contingency had been reinstated.

The present report describes an experiment undertaken in part-fulfilment of the regulations governing the degree of B.Sc. (Honours Psychology) in the Queen's University of Belfast, by S. D. Dalrymple.

References

Dufort, R. H., Guttman, N., and Kimble, G. A. (1954). One-trial discrimination reversal in the white rat. *J. Comp. physiol. Psychol.*, **47**, 248–9.

Levine, M. (1959). A model of hypothesis behavior in discrimination learning set. *Psychol. Rev.*, **66**, 353–66.

North, A. J. (1950*a*). Improvement in successive discrimination reversals. *J. comp. physiol. Psychol.*, **43**, 442–60.

North, A. J. (1950*b*). Performance during an extended series of discrimination reversals. *J. comp. physiol. Psychol.*, **43**, 461–70.

Pubols, B. H. (1957). Successive discrimination reversal learning in the white rat: A comparison of two procedures. *J. comp. physiol. Psychol.*, **50**, 319–22.

Pubols, B. H. (1962). An application of Levine's model for hypothesis behavior to serial reversal learning. *Psychol. Rev.*, **69**, 241–5.

Siegel, S. (1956). *Non-parametric Statistics for the Behavioral Sciences*. London: McGraw Hill.

Stretch, R., McGonigle, B., and Rodger, R. S. (1963). Serial position-reversal learning in the rat: A preliminary analysis of training criteria. *J. comp. physiol. Psychol.*, **56**, 719–22.

Stretch, R., McGonigle, B., and Morton, A. (1964). Serial position-reversal learning in the rat: Trials-per-problem and the inter-trial interval. *J. comp. physiol. Psychol.*, **57**, 461–3.

Manuscript received 20th December, 1965.

SPONTANEOUS SPEECH GENERATION

BY

EVERDINA A. LAWSON

*From the Department of Phonetics, University College, London**

A speech generation task was performed by bilingual subjects whilst they received irrelevant messages in one ear. The irrelevant messages varied in content as well as in language. Although these messages appeared to have a significant influence on the rate of speech, this variation was not consistent with any of the three hypotheses suggested. A further speech generation experiment was then carried out in which subjects received as irrelevant auditory input, in one ear, either a prose passage or "emotional" words repeated over a period of 1 min. Although the rate of speech did not seem to be affected by the irrelevant input, a memory test for words spoken, as distinct from words heard, seemed to indicate that the "emotional" words were significantly better recalled than the words from the prose passage. When a control experiment was performed with prose or repeated neutral words as auditory input, no such difference in recall was obtained. This result was seen as favouring Deutsch's model of the blocking of irrelevant speech.

Introduction

In a model proposed by Treisman (1961) for the perception of speech, it was suggested that part of the verbal input became selected for transmission whilst the remainder of the input became excluded through gradual attenuation or blocking. This selection takes place, through the operation of different procedures which make use of physical characteristics, transition probabilities, and past frequency of recognition and emission of particular words, at that level in the nervous system where the dictionary information is represented.

Deutsch and Deutsch (1963) suggest that, rather than a gradual blocking or attenuation of some of the input, all the available stimuli are transmitted to the dictionary units. Selection takes place for storage or output purposes on the basis of a preset strength of the dictionary units which is determined by the temporary or permanent importance of these units. In this view the same attenuating factors can be assumed to operate, but these would have their impact all at once rather than gradually.

It was thought that some evidence in favour of either model could be obtained with an experiment in which the subjects had to generate speech, whilst at the same time they were subjected to an input of irrelevant speech. If this irrelevant speech reached the dictionary units then it would seem possible that the activated units would be utilized for the speech sample on those occasions where this was appropriate. This could result in greater verbal fluency which could possibly be measured by counting the numbers of words generated per minute. Alternatively, the rate could slow down as a result of a decreased capacity to process information for output due to the amount of information transmitted with the input. If on the other hand the irrelevant speech was subjected to a gradual attenuation process then it would seem that again the rate of speech ought to increase as a result of an effort to shut this speech out more effectively, as suggested by Treisman (1961).

Although it was realized that difficulties in the interpretation of the results might occur since an increased speaking rate could be seen as evidence for different hypotheses, at least a decreased rate of speech would be evidence for one of the two models.

Experimental Procedure

The subject was instructed that his task was to talk continuously, at his usual speaking rate, for a period of 1 min. He could do so on any topic he chose. If the subject thought himself unable to do so for 12 periods of 1 min., he was given one of the several cartoons by Giles to study. After he said he had grasped the point of the cartoon, he was asked to talk about it. The cartoons were presented in random order. After the subject had commenced talking, the tape recorder was started, and he received, in one earphone only, the irrelevant messages in the following sequence:—

* Now at Department of Social Sciences, Institute of Technology, Bradford.

A1. 100 words of prose from "Lucky Jim."
A2. 100 words of prose from a Dutch novel.
A3. 100 words 8th order English approximations.
A4. 100 words 2nd order English approximations.
A5. 100 words 8th order Dutch approximations.
A6. 100 words 2nd order Dutch approximations.
B1. 100 words of prose from a Dutch novel.
B2. 100 words of prose from "Lucky Jim."
B3. 100 words 8th order Dutch approximations.
B4. 100 words 2nd order Dutch approximations.
B5. 100 words 8th order English approximations.
B6. 100 words 2nd order English approximations.

The English approximations were taken from Taylor and Moray (1960) and their method was used by the present author to obtain the relevant Dutch approximations. Three bi-lingual subjects (Dutch-English) were used for this experiment. The experiment was performed in two languages since it was designed as a consequence of an (unpublished) experiment on simultaneous translation.

Results

None of the subjects showed any intrusions from the auditory input, neither had any subject been aware of the language or the type of prose presented in the auditory message. Therefore, the number of words generated per minute were counted to see whether any influence from the auditory message could be detected. S1 and S3 made use of the cartoons, whilst S2 managed to talk spontaneously for the 12 1-min. sessions (see Table I and Fig. 1).

TABLE I

Number of Words Spoken per Minute

Subject	*S1*	*S2*	*S3*	*Mean*
Dutch spoken:				
A1 (rej. message English Prose) ..	150	222	171	181
A2 (do. Dutch Prose) ..	152	205	171	176
A3 (do. 8th order English) ..	148	188	148	161
A4 (do. 2nd order English) ..	172	212	165	183
A5 (do. 8th order Dutch) ..	166	182	155	167
A6 (do. 2nd order Dutch) ..	163	202	151	172
English spoken:				
B1 (rej. message Dutch prose) ..	142	186	162	163
B2 (do. English prose) ..	176	193	163	177
B3 (do. 8th order Dutch) ..	128	181	158	156
B4 (do. 2nd order Dutch) ..	172	206	182	187
B5 (do. 8th order English) ..	141	192	177	170
B6 (do. 2nd order English) ..	140	209	168	172

TABLE II

Source	*Sum of sq.*	*d.f.*	*Mean sq.*	*F*	*P*
1. Subjects	12,752	2	—	—	—
2. A v. B	61	1	61	0·48	>0·05
3. Conditions 1–6 ..	2,272	5	454·4	3·56	<0·025
Interaction 2 × 3 ..	485	5	97	0·76	>0·05
Residual	2,806	22	127·5		
	18,376	35			

(Residual containing interaction 1 × 3 and 1 × 2.)

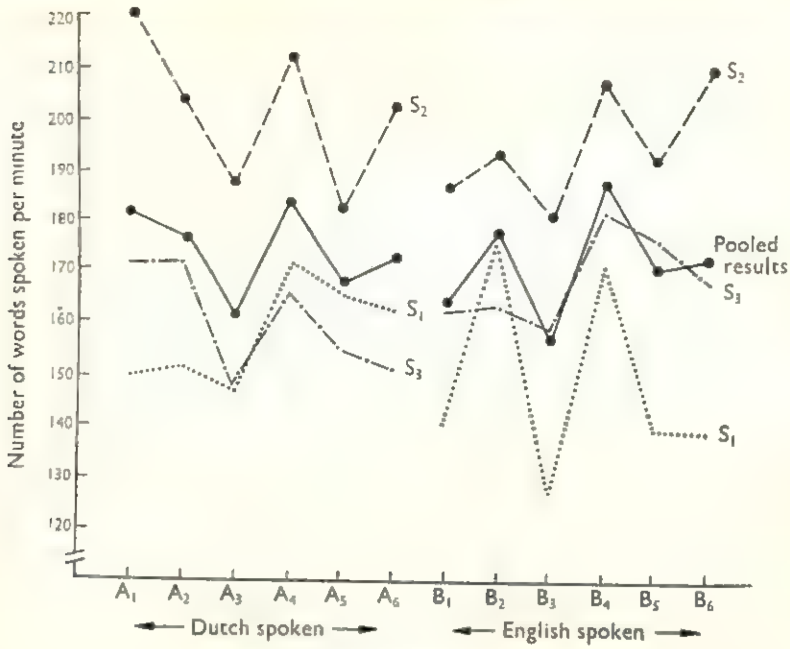


FIGURE 1

A three-way analysis of variance was carried out (see Table II).

The different conditions A1–6 and B1–6 cause a variance estimate which would be expected less than 25 times in 1,000 as a result of random variation.

Discussion

The analysis of variance indicates that the different conditions under which the speech was generated result in varying rates of speech.

Although the patterns of these speaking rates appear to be similar for the three subjects, it is not clear whether this is meaningful in terms of one of the three hypotheses.

If the rejected speech was in any way facilitating the generation process then it would be expected that prose in the same language would show this property most, since the same vocabulary would be utilized, and high transition probabilities would be present. That is, a peak should obtain at A2 (B2) and a lesser one at A5 (B5) which would slope down slightly to A6 (B6). A visual inspection of the graphs provides little encouragement for the maintenance of this assumption. If on the other hand the input speech diminishes the capacity available for the generating process, then the slowest rate ought to obtain when the rejected speech consists of passages with low transition probabilities, i.e. where the information content was highest, that is in A4 (B4) and A6 (B6). Again the graphs do not show any evidence in favour of this assumption. The third hypothesis of increased rate when a greater effort is needed to counteract interference of the rejected message, would predict the highest rate at A2 (B2) and the lowest rate at A4 (B4), since at A2 prose of the same language would be most interfering, and at A4, 2nd order passages in a different language would be least interfering (Broadbent, 1958, the most interfering materials are those most similar to the actual stimuli). Here the assumption is made that the speech generating process is subject to similar limitations as the reception of auditory stimulation but this is not necessarily the case.

Although it seems difficult to find a satisfactory interpretation of the results obtained, it does appear that the auditory input exerts some influence on the speech generating process (in view of the result of the analysis of variance). A further experiment was designed to subject this possibility to another test.

Experiment 2A

It was thought that if the subjects were unaware of the related message during shadowing, then the same should apply during speech generation, particularly in view of the fact that in the previous experiment the subjects could not report anything from the message transmitted to one ear. If however these words have a lowered threshold because of their emotional quality, and the dictionary in which these words had been activated was also required for the speech production process, then these words could possibly occur in the speech output. But if this were not the case then it could still be that evidence for the words having reached the dictionary units could be obtained from a memory test.

Again, some of Giles' cartoons were used to serve as material for the speech generation process.

(1) For the first minute the subject had to speak, he received in one ear a 1-min. passage from "Lucky Jim."

(2) For the second minute of speech generation, the subject received in one ear the following sequence of words: "tender and lovely and happy to cherish and cuddle and" repeated until the end of the 60 sec.

(3) For the third minute of speech generation the subject received in one ear the sequence: "Vicious and savage with fury and horror and anger and" repeated to the end of the minute.

Results

TABLE III

Number of Words Spoken

Subject	*Condition 1* (*"Lucky Jim"*)	*Condition 2* (*"Lovely," etc.*)	*Condition 3* (*"Vicious," etc.*)
S1	150	142	149
S2	136	128	150
S3	137	141	120
S4	75	79	100
S5	160	133	131
S6	144	155	126
Mean	134	130	129

The number of words produced per minute in each condition were to be counted, and a memory test for words heard or spoken was to be administered.

Six subjects, all undergraduate or postgraduate students from the psychology department, were used.

On a Friedman two-way analysis of variance the differences between the three conditions are not significant.

TABLE IV

Number of Words Recalled

Subject		*S1*	*S2*	*S3*	*S4*	*S5*	*S6*	*Total*
Condition 1.	Spoken words	3	4	3	4	4	2	20
	Auditory input	—	—	1	2	—	1	4
	Not occurring	—	—	1	1	1	—	3
Condition 2.	Spoken words	4	1	4	4	4	2	19
	Auditory input	4	3	4	4	—	3	18
	Not occurring	1	—	2	—	2	—	5
Condition 3.	Spoken words	2	3	4	3	3	3	18
	Auditory input	2	4	2	3	1	4	16
	Not occurring	—	—	—	—	—	—	—

For the memory test for words spoken and words heard, four words were selected from the generated speech, four from the auditory input and four which had not occurred at all. These were presented, in random order, to each subject immediately after he completed each condition of speech generation.

A Friedman two-way analysis of variance on the spoken words recalled in the three different conditions gives $\chi^2_r = 0{\cdot}7$ with $d.f. = 2$ which is not significant, i.e. the words recalled from the spoken words in the three conditions do not differ significantly.

The same test applied to the number of words recalled from the auditory input under the three conditions gives $\chi^2_r = 7{\cdot}58$, $d.f. = 2$ ($p < 0{\cdot}025$). Therefore the number of words recalled from the auditory input differ significantly for the three conditions. Condition 1 contributes most to this χ^2_r which seems the only significantly different condition since ΣRj for condition 2 and 3 are almost equal (9·5 and 9 respectively).

Since it appeared that the emotional words were better recalled than the neutral ones, a control experiment was performed to test whether the repetition and the consequent rhythm were responsible for the better recall.

Experiment 2B

The procedure was similar to that of experiment 2A, cartoons again being used when required. Two 1-min. passages were recorded for the auditory input, as in experiment 2A, i.e. in a monotone, and as much as possible, at a constant intensity and constant speed. One of the two passages was again taken from "Lucky Jim" and the second was the following sequence of words: "large and empty and small to sing and walk and," repeated until 1 min. was completed.

Six subjects were used, drawn from the same population as those for experiment 2A.

TABLE V

Number of Words Spoken in 1 min.

Subject	*Condition* 1	*Condition* 2
S1	168	158
S2	151	168
S3	124	139
S4	160	156
S5	123	127
S6	144	136
Mean	145	147·8

Wilcoxon—number of spoken words for condition 1 and 2 do not differ significantly for N = 6 (one tie).

TABLE VI

Number of Words Recalled

Subject		S1	S2	S3	S4	S5	S6	*Total*
Condition 1.	Spoken words	3	3	3	3	4	4	20
	Auditory input	—	—	—	—	1	—	1
	Not occurring	—	—	—	1	—	—	1
Condition 2.	Spoken words	4	4	4	4	3	4	23
	Auditory input	—	—	2	—	1	—	3
	Not occurring	—	—	—	—	—	—	—

Discussion

From these two experiments it appears that, although no words from the auditory input occurred in the generated speech, nevertheless the auditory input does not necessarily remain "shut out." It would seem in fact that the words with the emotional character are noticed, whilst the comparable neutral ones are not. This can hardly be attributed to the repetitive character of the auditory input in experiment 2A since the same does not obtain for the similarly repetitive auditory input of the neutral words in experiment 2B. Moreover, Moray (1960) repeated a short sequence of words in the rejected message of a shadowing experiment 35 times without this being noticed by his subjects. Furthermore, although shadowing and speech generation do not necessarily make use of the same mechanisms to the same extent, Treisman (1961) found that shadowing and speech generation were not significantly different with regard to the rejected message.

Therefore, it would appear that these "emotional" words have a quality which makes them easier to discriminate than comparable neutral words. This possibility had been suggested by Moray (1960) but has proved rather elusive to verify experimentally.

Although no words from the auditory input emerged in the generated speech, two conclusions could possibly be drawn from this experiment. Firstly, that although the task of speech generation requires, like shadowing, most of the available attentional capacity, nevertheless other perceptual input can be noticed, providing the threshold of the particular input stimuli is low enough.

Secondly, the results obtained would seem to favour an assumption of one dictionary system for input and output. Were this not the case, but instead a different dictionary used for the output, then it seems most unlikely that anything at all from the auditory input would be noticed, considering the very limited amount of attentional capacity available for this irrelevant input. That the attentional capacity was limited would seem apparent from the fact that no other words but the emotional ones were recalled. Although the information capacity of the spontaneous speech is not known, it seems reasonable to refer back to Treisman's (1961) experiments and to take it that no significant difference regarding the rejected message between shadowing and speech generation relates to available capacity at the dictionary level. If the same dictionary was utilized, but the words from the auditory input could not be incorporated in the output because of possibly the lack of appropriate connections to the higher levels in the nervous system where the speech generating process was initiated, this could indicate that the blocking of the auditory input might in fact take place at a higher level than the dictionary system.

Although no conclusive evidence has emerged from these experiments in favour of Treisman's or Deutsch's model, the balance seems to be tilted in favour of the latter model. It seems improbable that an irrelevant verbal input which is almost completely attenuated before reaching the dictionary system nevertheless exerts a noticeable influence over word output selection. The differential results obtained from the speech generation task, where an increased or decreased rate of word output was obtained as a result of different irrelevant messages in one ear, appear to suggest that these irrelevant words actually reached the dictionary. Although no unequivocal explanation of this occurred, the finding seems more compatible with Deutsch's theory.

These experiments formed part of a thesis submitted for the Ph.D. degree of London University. My thanks are due to Professor D. Fry and Dr. F. Goldman-Eisler of University College, London, for support and supervision of the research. Also to D.S.I.R. for financial support as well as to the Central Research Fund, whose grant made the acquisition of the required apparatus possible.

References

Broadbent, D. E. (1958). *Perception and Communication.* London: Pergamon.

Deutsch, J. A., and Deutsch, O. (1963). Attention: some theoretical considerations. *Psychol. Rev.*, **70**, 80–90.

Moray, N. (1960). *Selective Listening.* Unpublished D.Phil. thesis, Oxford.

Taylor, A. M., and Moray, N. (1960). Statistical approximations to English and French. *Language and Speech*, **3**, 7–10.

Treisman, A. M. (1961). *Attention and Speech.* Unpublished D.Phil. thesis, Oxford.

Manuscript received 8th November, 1965

DECISIONS CONCERNING THE REJECTED CHANNEL

BY

EVERDINA A. LAWSON*

From the Phonetics Department, University College, London

It was thought that the physical aspects of auditory stimuli were possibly transmitted via separate pathways from those transmitting the verbal aspects. Three experiments were designed to test this hypothesis. In these experiments subjects had to perform a shadowing task and had to respond simultaneously on response keys to pips superimposed in either ear on verbal messages. The response to these pips was of increasing complexity, in that it was a simple reaction time which was measured in the first experiment, a choice reaction time in the second experiment and a more complex choice reaction time in the third experiment. Subjects were able to perform these tests although the increasing difficulty was reflected in longer reaction times and more errors. The reaction times to the pips presented to the ear which was not being shadowed were slower, and the errors, made to pips in both channels, were "false positives" rather than errors of omission. These results were taken as favouring the hypothesis.

Introduction

In experiments on the transmission of verbal stimuli it has been assumed that such stimuli arrived, at some stage in the nervous system, at a place where the elements of a dictionary are represented (Treisman, 1961; Deutsch and Deutsch, 1963). Experimental results obtained by Treisman (1961) and others seem to indicate that the physical properties of a sequence of auditory stimuli can be detected, whilst the verbal content of such a message remains unnoticed. The continued availability of the physical characteristics, such as voice and location of a message and the apparent blocking of the verbal content, could be the result of separate pathways for transmission of these aspects of auditory stimuli rather than the coincidence or sequence of such pathways. It was thought that experimental evidence in favour of an assumption of separate pathways could be obtained by presenting subjects with a shadowing task, whilst at the same time requiring them to handle simultaneous physical stimuli. Three experiments were designed in which subjects had to shadow and respond to pips in either ear by pressing a response key.

Procedure

Experiment 1

A 10-min. passage from "Lucky Jim" was recorded on one channel of an E.M.I. tape recorder, and an equally long passage from "Dr. Zhivago" was recorded on the other channel. On each of the two channels of a Stella tape recorder, 20 "pips" of 750 cps. of 350 millisec. duration were recorded. These "pips" were randomly distributed over the two channels, each channel containing 10 pips, and were at an interval of 13–18 sec., these intervals again being randomly distributed.

The subject was instructed to shadow the "Lucky Jim" passage and at the same time to respond with his left hand on the appropriate key to pips in his left ear and with his right hand on the other key to pips in his right ear. This response had to be made as quickly as possible without however impairing the shadowing performance, since this was the task of major importance. After 5 min. shadowing, the experiment was interrupted and the subject turned his earphones round and shadowed the same passage on the other ear. This was to counteract any difference in ear sensitivity. The order in which the ears were used was randomized over the subjects. The pips from the Stella tape recorder were connected with the mixing unit and from there to the earphones, as well as via a modifying circuit to the chronotron. In this way, the pips commenced the time recording on the chronotron whilst the subject terminated this by pressing his response key. Six subjects were used.

* Now at the Department of Social Sciences, Institute of Technology, Bradford.

Experiment 2

Three further pairs of 10-min. passages were recorded on the E.M.I. recorder. Again the passages to be shadowed were taken from "Lucky Jim" and the passages for the rejected ear from "Dr. Zhivago." The pips recorded on the Stella recorder were of the same duration as those in the previous experiment, but with the pitch varied, in that two tones were recorded, one of 1,000 cps. and one of 750 cps. Each tone was presented 10 times in each channel for the first pair of passages. They were presented in random order and at intervals, randomly varied from 13–18 sec. The pips for the choice RT for the second set of passages were recorded on one channel of the Stella only, in order to compare RTs to pips in either channel with choice RTs to pips in one channel only. Each tone occurred 20 times during this passage. The pips were presented to either the shadowed or the rejected channel by switching the connections from the Stella recorder at the mixing unit. This switching of the connections of the pip transmitting channels was again done after 5 min. shadowing. The third passage was mixed with pips of the 750 cps. tone on one channel only, i.e. a simple reaction time being measured. Again the pip transmitting channels were switched half way through the passage.

The instructions to the subject were to shadow one ear and at the same time to respond to the lower pitched of the two pips presented with the right hand on the corresponding key for pips in the right ear, and similarly with the left hand for pips in the left ear. Differences in performance due to differences in ear sensitivity were eliminated by switching the earphones half way through the experiment. The order of ear shadowed first was again randomized. Six subjects were used.

Experiment 3

This experiment was similar to the previous one, in that again three 10-min. passages were recorded from "Lucky Jim" and "Dr. Zhivago" and on each passage 10 pips of 750 cps. and 10 pips of 1,000 cps. were superimposed. The instructions to the subject were also similar but were complicated in that the left hand response had to be only to the higher pitched pips which occurred in the left ear, and the right hand response only to the lower pitched pips which occurred in the right ear. For the third passage the subject did not shadow at all but he had to respond to the pips similarly as above. Eight subjects were used in this experiment.

Results

Experiment 1

TABLE I

Subject	*Mean RT in millisec. to pips in shadowed channel*		*Mean RT in millisec. to pips in rejected channel*	
	A *Left ear*	*B* *Right ear*	*C* *Right ear*	*D* *Left ear*
1	418	432	448	402
2	622·5	629	654	604·5
3	531	527	545	615
4	411	505	465	560
5	638	668·5	620	619
6	548	706	605	712
Mean	528	577·9	556·2	585·5

The mean RTs are shown in Table I. A three-way analysis of variance was carried out on the scores for each subject. The results of this are shown in Table II.

No significant differences resulted from either responses given to clicks in shadowed *versus* rejected ear, or from right *versus* left ear, but the interaction of shadowed/rejected channel and right/left ear, gives an estimate of the variance which is significant at the 0·05 level. The interaction subjects/channel, and subject/ears, was very small, and it appeared that a better test of the other variance estimates would result by taking these interactions into the residual. The shadowing response did not show any discontinuities.

TABLE II

Source	*Sum of square*	*d.f.*	*Mean square*	*F*	*p*
1. Subjects	164,027	5	—	—	—
2. Shadow/rejected	1,908	1	1,908	1·01	>0·05
3. Ears	640	1	640	0·34	>0·05
Interaction 2 × 3 ..	9,362	1	9,362	4·97	<0·05
Residual	28,264	15	1,884		
Total ..	204,201	23			

Experiment 2

TABLE III

	Subject	*Mean choice RT in millisec. to pips*			
		Shadowed ear	*Mistakes*	*Rejected ear*	*Mistakes*
A. Either ear ..	1	584	—	636	—
	2	592	—	745	3
	3	785	1	827	2
	4	720	—	856	3
	5	567	—	617	4
	6	725	—	790	2
	Mean	662·2	—	745·2	14/60 = 23 per cent.
B. One ear only ..	1	572	—	634	—
	2	772	—	950	2
	3	860	—	1048	2
	4	600	—	700	1
	5	570	—	640	2
	6	674	—	745	1
	Mean	674·7	—	786·2	8/60 = 13 per cent.
		Mean simple RT in millisec. to pips			
C. One ear only ..	1	542	—	698	—
	2	762	—	962	—
	3	840	—	895	—
	4	370	—	410	—
	5	520	—	609	—
	6	540	—	560	—
	Mean	595·7	—	689·0	—

The mean choice RTs for the first two passages and simple RTs for the third passage are shown in Table III. A three-way analysis of variance was carried out on these results (see Table IV).

TABLE IV

Source	*Sum of square*	*d.f.*	*Mean square*	*F*	*p*
			—	—	—
1. Subject	317,758	5	82,848	6·05	<0.025
2. Shadow/rejected	82,848	1	24,472	1·79	>0.05
3. Tasks A/B/C ..	48,944	2	10,624·5	0·77	>0.05
Interaction 2 × 3 ..	21,249	2	13,676		
Residual	341,903	25			
	812,702	35			

The choice-reaction times obtained from the channel which the subject is shadowing, differ significantly from those obtained from the channel to which the subject does not attend. (Interaction subject/task and subject/channel again taken into residual). Task A–B and C did not cause a variance which would require rejection of the null hypothesis.

All the mistakes made were responses to the wrong pip, i.e. were extra responses. Of the pips which required a response, none was missed. The shadowing response did not show any obvious discontinuity as a result of this double task.

TABLE V

Experiment 3

	Subject	*Mean choice RT in millisec. to pips in either ear*			
		Left hand	*Mistakes (per cent.)*	*Right Hand*	*Mistakes (per cent.)*
A. Left ear shadowed	1	603	14	684	40
	2	594	60	712	40
	3	1,160	40	1,200	40
	4	759	40	950	40
	5	669	40	705	40
	6	790	40	855	60
	7	835	40	910	40
	8	940	40	1,046	40
	Mean	793·7		882·7	
B. Right ear shadowed	1	595	40	526	20
	2	646	0	609	40
	3	1,350	20	1,250	40
	4	721	0	813	40
	5	740	60	685	40
	6	870	40	815	40
	7	896	40	858	40
	8	1,220	40	1,150	20
	Mean	879·7		838·2	
C. No shadowing ..	1	665	0	635	0
	2	696	20	713	0
	3	1,150	0	1,180	0
	4	945	0	961	0
	5	760	0	785	0
	6	790	0	775	0
	7	915	0	932	0
	8	822	0	780	0
	Mean	842·9		845·1	

The results of the third experiment are shown in Table V and the analysis of variance in Table VI. Interaction subjects/task was not taken into the residual because of its size. This interaction seems to indicate that there were differences between individual subjects in the performance of the different tasks.

Interaction ears/task seems the only indication that there was a difference between shadow/rejected ear, in responding to the specific pip.

TABLE VI

Source	Sum of square	d.f.	Mean square	F	p
1. Subjects	1,570,143	7	—	—	—
2. Ears	3,300	1	3,300	2·1	$>0·05$
3. Tasks A/B/C ..	3,673	2	1,836·5	1·2	$>0·05$
4. Interaction 2 × 3	35,293	2	17,646	11·4	$<0·001$
Interaction 1 × 3	231,064	14	16,504·8	10·7	$<0·001$
Residual	32,336	21	1,539·8		
Total ..	1,875,809	47			

Discussion

The only significant difference obtained in the first experiment was the variance estimate resulting from the interaction of task (shadowing or not shadowing), with right or left ear. The explanation of this could be that there was a slight difference of intensity between the pips, those transmitted via the pre-amplifier of the Stella recorder being slightly less loud. This difference was only realized after this significant interaction had been obtained.

The results of this first experiment seem to indicate that the transmission channels for these physical stimuli are monitored continuously since not only are the reaction times not significantly different, but also none of the pips in either channel were missed by any of the subjects. That this did not apply to the channels transmitting the verbal input would seem to follow from Treisman's (1961) different results which she obtained when inserting digits or the word "tap" in both the shadowed and the rejected message, as both digits and the word "tap" belong to the category of verbal input.

From the results of the second experiment it appears that even this more complex task can be performed reasonably satisfactorily. Apparently, shadowing is not the kind of task which occupies the information transmission channels to the extent that no other information, irrelevant to the shadowing, can be dealt with. However, Moray (1960) and Treisman (1961) have shown that the word transmission system is fully occupied in that almost no verbal information from the rejected channel was selected for attention at all. This combined evidence would seem to favour an assumption of separate channels for the transmission of words and for the transmission of purely physical stimuli.

From this experiment it would appear, however, that the ability to cope with the task with the ear which also receives the verbally rejected messages, is impaired, in comparison to the performance at the other ear. Not only are the reaction times longer at the rejected ear but also several mistakes were made. The mistakes were all false positives, i.e. reactions to the wrong pip, and none of the correct pips were missed.

That the simple reaction times in this experiment are longer than those in the first experiment, although the subjects were drawn from the same population, is probably due to the subjects' ignorance of the changed conditions. The subjects were not informed of the change and therefore were probably still acting upon a larger ensemble from which to select the stimulus, and were thus reacting more slowly. This would also seem to apply to the choice reaction times to the pips in one ear as compared with choice reaction times to pips in both ears, as the subjects had not been informed of the changed condition.

The third experiment also shows a significant difference between the responses to pips in the shadowed ear and those in the rejected ear, in that the variance estimate based on the interaction of ears and task is significant at the 0·001 level. The increased complexity of this task is apparent from the longer reaction times as well as from the considerably increased percentages of errors. As in Experiment 2, these errors were all

extra responses, i.e. responses to the wrong pip over and above those to the correct pips. The responses to the wrong pips were the only occasions where at times this task interfered with the shadowing task. This interference was generally a momentary confusion due to the apparent realization of a wrong response. Realization that a response was wrong occurred almost every time it was made and there was a tendency to verbalize the error. This finding is perhaps explained by Broadbent's (1958, p. 51) discussion of an experiment by Bates where the suggestion was made of an appropriate verbal response with an inappropriate bodily one.

In this experiment the errors are made in both channels and occur in each with almost equal frequency. This higher incidence of errors, added to the increased reaction times, seems consistent with the expectation that a breakdown in responses to these physical stimuli would have to occur, if these stimuli demanded an attentional capacity exceeding the remaining capacity which was not required for shadowing.

The results of these experiments seem to favour the assumption of separate pathways for physical stimuli. The experiment by Ladefoged and Broadbent (1960) would also seem to corroborate the present evidence. Presumably the channels transmitting the physical stimuli lead to a decision mechanism which operates on that input as well as on the transmission of the verbal contents of messages. Although the transmission of the verbal characteristics of a message is limited by the capacity of the dictionary system, this apparently does not imply that the attentional capacity is entirely exhausted by this shadowing task. Moreover, although a breakdown in dealing with a task consisting of physical auditory stimuli might occur, such an impaired performance does not seem to result in attenuation or blocking, as is the case when the verbal channels become overloaded, but rather in increased choice reaction times and increased number of errors.

These experiments formed part of a thesis submitted for the Ph.D. degree in the University of London. I am indebted to Professor Fry and Dr. Goldman-Eisler of University College, London, for supervision and general encouragement during the research. I am also grateful to D.S.I.R. for financial support during the research period as well as to the Central Research Fund for a grant which made the purchase of the necessary equipment possible.

References

Broadbent, D. E. (1958). *Perception and Communication*. Oxford: Pergamon.

Ladefoged, P., and Broadbent, D. E. (1960). Perception of sequence in auditory events. *Quart. J. exp. Psychol.*, **12**, 162–170.

Deutsch, J. A., and Deutsch, O. (1963). Attention, some theoretical considerations. *Psychol. Rev.*, **70**, 80–90.

Moray, N. (1960). *Selective Listening*. D.Phil. Thesis, Oxford.

Treisman, A. M. (1961). *Attention of Speech*. D.Phil. Thesis, Oxford.

Manuscript received 6th December, 1965.

THE ROLE OF THE INTERPOLATED TASK IN SHORT-TERM RETENTION

BY

R. CONRAD and A. J. HULL

From the Medical Research Council's Applied Psychology Research Unit, Cambridge

It has been proposed that a single set of operations based on classical interference theory is adequate to describe the phenomena of both short- and long-term memory. An article by Keppel and Underwood (1962) argues that short-term forgetting is due to proactive interference and, by implication, not a result of trace decay. An experiment which varied retention interval and the nature of the interpolated task, gave results which indicate that when the amount forgotten and the nature of errors are considered, a decay model is supported, the proactive interference suggestion being untenable.

Introduction

A crucial question in the field of verbal learning theory which is currently generating discussion, is whether or not it is useful to regard short-term memory (STM) and long-term memory (LTM) as being supported by the same mechanisms. In general, authors writing from the standpoint of classical interference theory (Keppel and Underwood, 1962; Melton, 1963) have favoured a single set of operations to govern retention over both short and long time intervals. But authors proposing models for STM which essentially involve information loss during storage (e.g. Broadbent, 1963; Waugh and Norman, 1965) have argued that this storage system must be independent of that required for storage over long time intervals.

The latter type models always include an operation which permits a fading trace in the short-term store to be rescued up to a certain point in its history, and transferred into the more permanent long-term store. The necessary condition for this eventuality is rehearsal of some kind; or, perhaps another way of saying the same thing, recoding of the input with increased redundancy (Brown, 1959). If conditions during the retention interval preclude rehearsal, then after a short time the trace will become irrecoverable. Recoding or rehearsal is held to modify the trace so that it takes on a relatively permanent status. Rapid decay occurs in one state (store) but not in the other.

On the other hand, Keppel and Underwood (1962) have proposed that interference theory, using a single set of assumptions, can account for forgetting regardless of the duration of the retention interval. Keppel and Underwood have examined in some detail the situation where a subject is required to retain a letter- or digit-sequence well within the memory span, during an interval filled with an activity presumed (*a*) to be non-interfering and so ruling out retroactive inhibition, and (*b*) to preclude rehearsal. The argument is that when, as is usual, multi-trial tests are given, ideal conditions are present for the occurrence of proactive interference (PI) from previous trials. For example, trial-2 learning is held to extinguish trial-1 associations, which recover during the retention interval to compete at recall. The longer the retention interval, the greater the recovery and the poorer therefore the recall of trial-2 items. How valid is this?

As well as the demonstrated relation between duration of retention interval and recall, certain other predictions would seem to follow from Keppel and Underwood's argument. Firstly, as retention interval increased and trial-1 associations continued to recover in strength, trial-2 recall errors should show an increasing proportion of serial position intrusions from trial-1. Peterson and Gentile (1965) however reported a relative decrease in such intrusions with increased retention interval. Secondly when Conrad (1966) analysed recall errors from consonant quadrigrams, he showed that as retention interval increased, errors became increasingly random and less and less acoustically similar to the correct consonant. He argued that this could only be predicted from interference theory if the assumption were made that degree of extinction depended on dissimilarity between trials-1 and -2 items. This is the opposite to the reasonable prediction. In view of the failure of two predictions the present experiment tests the critical assumption that degree of recovery depends on the duration of the retention interval.

Method

Procedure

With minor refinements, the procedure was that described by Brown (1958) and Peterson and Peterson (1959). Using a punched-tape programme, a Bina-View display device presented, one at a time, four consonants followed by a series of digits. Subjects were instructed to read all items aloud but to report only the consonant sequence in correct order on a prepared answer sheet. The presentation rate was controlled by electronic timer, the cessation of clicks being the signal for recall, and the rate for consonants and digits could be independently varied. Reading the digits served merely as a non-interfering filler during the retention interval. There was a fixed interval of 15 sec. between the last digit of one trial and the first of three warning blips of the next trial.

Materials

The quadrigrams were selected from a vocabulary of 10 consonants, which were printed in alphabetic order on the subject's answer sheet. The selection served merely to provide test material of difficulty known from previous experiments. Four tests each of 10 quadrigrams were prepared, such that the four letters of each quadrigram were different from each other, and different from those of the immediately preceding quadrigram; and each letter occurred equally often in each serial position.

Design

There were four experimental conditions, viz.:

8F	—quadrigram at 0·4 sec./letter	+ 8 digits at 0·4 sec./digit
8S	,, ,, 0·4 ,,	+ 8 ,, ,, 0·8 ,,
16F	,, ,, 0·4 ,,	+16 ,, ,, 0·4 ,,
16S	,, ,, 0·4 ,,	+16 ,, ,, 0·8 ,,

Each condition was represented by a test of 10 trials. Twenty-four subjects, some naval ratings, some housewives, did all conditions in a completely counterbalanced order. Each condition used the same consonants, and by previous experiment, tests had been equated for recall difficulty. For each subject testing was completed within a single session. The session began with four practice trials, and the 24 subjects used were all able to follow the instructions.

Results and Discussion

Performance was scored firstly for mean percentage of letters wrong taking serial position into account, and secondly for percentage of errors that were serial position intrusions (SPI's), i.e. when the wrong letter repeated the letter given (correctly or otherwise) in the same position of the immediately preceding quadrigram. These values are shown in Table I.

TABLE I

Effect on Recall of Interpolated Task and Retention Interval

	Conditions			
	8F	*8S*	*16F*	*16S*
Per cent. wrong letters	35·8	30·3	43·1	32·9
Per cent. serial position intrusions	22·9	19·0	15·1	14·8

On wrong letters, analysis of variance shows significant effects of the number of interpolated digits ($p < 0.001$), the rate of reading digits ($p < 0.001$) and subjects ($p < 0.01$). The number × rate interaction does not quite reach significance at the 0·05 level (F (1·66) = 3·59). The significance of the difference between conditions using

the criterion of wrong letters was tested by studentized range statistic (Winer, 1962), with the following results:

$$\begin{array}{r}
8F\ v\ 16S—p > 0{\cdot}05 \\
8F\ v\ \ 8S—p < 0{\cdot}05 \\
16F\ v\ 16S—p < 0{\cdot}01 \\
8F\ v\ 16F—p < 0{\cdot}01 \\
8S\ v\ 16S—p > 0{\cdot}05 \\
8S\ v\ 16F—p < 0{\cdot}01
\end{array}$$

From the immediate point of view the comparison 8S *v* 16F is the most important. It shows that duration of retention interval is not in itself crucial. Rather, when the retention interval is unchanged, events during the interval are the significant determinants of recall.

The data on serial position intrusions, though not entirely conclusive, provide little comfort for the recovery assumption. The Keppel and Underwood (1962) argument would lead to the prediction that increased retention interval would lead to increased proportion of such errors. Of the five comparisons between pairs of conditions where retention interval is different, all in fact, show the opposite effect. The longer the interval the smaller the proportion of intrusions—indeed they are fewer absolutely. But only two comparisons (8F *v* 16F; 8F *v* 16S) show a significant difference at better than the 0·05 level. Where the retention interval is the same (8S *v* 16F) the difference is not significant ($0{\cdot}05 < p < 0{\cdot}1$).

The wrong letter results of the first five comparisons can be equally well predicted on an assumption of recovery of traces during retention, or of decay during retention. In both cases recall would be expected to decline with the longer interval, either because greater recovery would lead to increased competition at recall, or because traces would have longer to decay. The last comparison though (8S *v* 16F) where retention interval is the same gives a result which is expected neither on a simple recovery, nor on a simple time-decay basis. But this is of course the crucial comparison which involves constant retention interval but different events during that interval.

The effect of increasing shadowing rate would in decay theory terms reduce the opportunity for rehearsal (Broadbent, 1963; Pollack, 1963; Posner and Rossman, 1965). Granted that, the direction of the observed difference between 8S and 16F is what would be expected. But neither would the assumptions of interference theory be violated if shadowing rate were held to affect degree of learning as measured at the end of the retention interval: this also would be to the advantage of 8S. So at first sight, the aberrant result along with the others could be taken to be in accord with both theories. But recourse to the rehearsal or degree of learning argument has implications concerning SPI errors which appear to be crucial to the discussion.

With degree of learning constant, recovery time (i.e. retention interval) would reflect in probability of SPI's. Keppel and Underwood (1962, p. 160) are clear that: "Such intrusions . . . represent the evidence needed to support the notion of spontaneous recovery of extinguished or partially extinguished associations over short intervals." If degree of learning is constant, the longer the recovery time, the greater the proportion of errors which would be SPI's. Similarly, with recovery time constant faster shadowing rate (less degree of learning) would also lead to proportionately more SPI's. If both vary, no prediction can be made. Conditions for prediction therefore exist in the case of three comparisons: 8S *v* 16S, 8F *v* 16F, 8S *v* 16F. Respectively as a proportion of errors, there should be more SPI's for 16S, 16F and 16F. Table I shows that the prediction fails in all three cases. In the first and last case the difference (in the opposite direction to that predicted) does not reach significance at the 0·05 level ($0{\cdot}05 < p < 0{\cdot}1$); in the second case the difference which is in the wrong direction is highly significant ($p < 0{\cdot}001$).

On almost any decay model, and explicitly on that suggested by Conrad (1965), in this situation, during the retention interval rehearsal is fighting a losing battle against decay. But the rehearsal could only be concerned with the items of the current trial. The items of the previous trial which would be involved in SPI's, would simply be decaying. The probability of SPI therefore depends only on time between recall of trial-1 items and recall of trial-2 items (Conrad, 1960; Peterson and Gentile, 1965). Since the inter-trial interval was constant, retention interval is the crucial one. When pairs of conditions are compared, regardless of shadowing rate, more SPI's are expected for the condition with the shorter interval. The Table I data show this to hold in every case. In the 8S *v* 16F comparison where retention interval is the same, the difference is not significant.

Conclusions

The assumption of recovery of extinguished associations during subsequent retention is clearly in doubt. In its present form, the Keppel and Underwood PI model seems inadequate to account for the data. As those authors state it, the effects of varying shadowing rate (or other interpolated tasks of differing complexity) are not accounted for. The SPI data are counter to the PI predictions even if the degree of learning assumption is added. Wickelgren (1965) has attempted to account for the results of typical short-term memory experiments in terms of retroactive interferance. But there is not the slightest evidence in the present experiment, nor in any others, that interpolated digits interfere, in the classical sense, with learned letters. At present therefore, decay models seem better able to handle forgetting over these short time intervals; in which case without doubt, theoretical amplification is badly needed to cover the ground between retention intervals of seconds, and those of days.

References

BROADBENT, D. E. (1963). Flow of information within the organism. *J. verb. Learn. verb. Behav.*, **2**, 34–9.

BROWN, J. (1958). Some tests of the decay theory of immediate memory. *Quart. J. exp. Psychol.*, **10**, 12–21.

BROWN, J. (1959). Information, redundancy and decay of the memory trace. In *The Mechanization of Thought Processes*. Nat. Phys. Lab. Symp., No. 10, H.M.S.O.

CONRAD, R. (1960). Serial order intrusions in immediate memory. *Brit. J. Psychol.*, **51**, 45–8.

CONRAD, R. (1965). Order error in immediate recall of sequences. *J. verb. Learn. verb. Behav.*, **4**, 161–9.

CONRAD, R. (1966). Interference or decay over short retention intervals. *J. verb. Learn. verb. Behav.* (in press).

KEPPEL, G., and UNDERWOOD, B. J. (1962). Proactive inhibition in short-term retention of single items. *J. verb. Learn. verb Behav.*, **1**, 153–61.

MELTON, A. W. (1963). Implications of short-term memory for a general theory of memory. *J. verb. Learn. verb. Behav.*, **2**, 1–21.

PETERSON, L. R., and GENTILE, A. (1965). Proactive interference as a function of time between tests. *J. exp. Psychol.*, **70**, 473–8.

PETERSON, L. R., and PETERSON, M. J. (1959). Short-term retention of individual verbal items. *J. exp. psychol.*, **58**, 193–8.

POLLACK, I. (1963). Interference, rehearsal and short-term retention of digits. *Canad. J. Psychol.*, **17**, 380–92.

POSNER, M. I., and ROSSMAN, E. (1965). The effect of size and location of informational transforms upon short-term retention. *J. exp. Psychol.* **70**, 496–505.

WAUGH, N. C., and NORMAN, D. A. (1965). Primary memory. *Psychol. Rev.*, **72**, 89–104.

WICKELGREN, W. A. (1965). Acoustic similarity and retroactive interference in short-term memory. *J. verb. Learn. verb. Behav.*, **4**, 53–61.

WINER, B. J. (1962). *Statistical Principles in Experimental Design*. New York: McGraw Hill.

Manuscript received 18th January, 1966.

ON DETERMINING PAIN THRESHOLDS USING THE LIMITING METHOD

BY

E. A. C. THOMAS

From the Department of Mathematical Statistics, University of Cambridge

A model developed by Cane (1956) for the study of relative thresholds is modified to apply to absolute thresholds. It is assumed that the fluctuation in the subject's sensation of a stimulus is uniformly distributed, and the model is used to estimate the threshold from data presented by several authors.

Introduction

In experiments to determine heat-pain thresholds the subject is exposed to stimuli of increasing magnitudes and is required to signal when a pricking pain first occurs immediately after a stimulus. This stopping point is then used to estimate the threshold. A statistical analysis of this so-called Limiting Method has shown that the distribution of the stopping point depends on the size of the stimulus increment used and, therefore, before the results of different threshold experiments are compared a suitable correction should be made (Brown and Cane, 1959). In a recent paper (Haslam, 1965), it is contended that there is a residual effect of stimulus scale-interval even when correction has been made for the statistical effect predicted by Brown and Cane. It is the purpose of this note to examine this contention.

A Model

The model developed by Cane (1956) for the study of relative thresholds is easily modified to apply to absolute thresholds. Let us assume that when the intensity of the stimulus is t units the subject's sensation is $t + x$ units, where x is a random variable with density function $f(x)$, and is independent of t. We assume further that there is a threshold, c, such that the subject experiences pain only if $t + x > c$. So that, if a stimulus of intensity t is presented, the probability, $p(t)$, that pain is experienced is given by

$$p(t) = \int_{c-t}^{\infty} f(x)dx. \qquad (1)$$

The absolute threshold, T^*, is then defined by the relation

$$p(T^*) = \alpha \text{ for some positive } \alpha < 1.$$

It will be assumed here that $\alpha = 1/2$, i.e. that T^* is the intensity which the subject says causes pain on 50 per cent. of the occasions. If $f(x)$ is symmetric, $p(c) = \int_0^{\infty} f(x)dx = 1/2$, so that in this case $c = T^*$.

Let $t_0 < t_1 < t_2 < \ldots$ be the values of t at successive presentations; let $q_r = 1 - p(t_r)$ for $r = 1, 2, 3, \ldots, \pi_0 = 1$, and $\pi_i = q_1 q_2 \ldots q_i$. Then the probability of stopping at t_i is $(\pi_{i-1} - \pi_i)$. If the steps are of equal size $2K/n$, the expected value of the stopping point, T, is given by

$$E(T - t_0) = \frac{2K}{n} \sum_{i=0}^{\infty} \pi_i \qquad (2)$$

$$\text{and } E((T - t_0)^2) = \frac{4K^2}{n^2} \left[\sum_{i=0}^{\infty} \pi_i + 2 \sum_{i=1}^{\infty} i\pi_i \right] \qquad (3)$$

To estimate T^* we have to make some assumptions about $f(x)$ (or, correspondingly, about $p(t)$) and then use the experimental results to estimate any unknown parameters. Brown and Cane (1959) have shown that if $p(t)$ is cumulative normal with standard deviation σ, and t_0 is chosen so that $p(t_0) \leqslant \frac{1}{2\pi} \int_{-\infty}^{-3} exp\,(-x^2/2)\ dx$, $K = 3\sigma$, then T^*

may be estimated by

$$\bar{T} - \left(\frac{4{\cdot}86}{\sqrt{n}} - 1{\cdot}67\right)\sigma \quad \ldots \quad (4)$$

where $\bar{T}$ is the mean stopping point for the subject(s). σ^2 is the variance of the random variable, x, and is, in general, not equal to the experimental variance, $var(T)$, for the Limiting Method, though it may be estimated from it. This point seems to have been overlooked by Haslam (1965).

If, as a first approximation, we assume that

$$\begin{aligned} p(t) &= 0 && \text{for } t < t_o \\ &= t - t_o/2K && t_o \leqslant t \leqslant 2K + t_o \\ &= 1 && t_o + 2K \leqslant t \end{aligned} \quad \ldots \quad (5)$$

then $T^* = t_o + K$. Let n be the integral part of $2K/s$, s being the step-size. Then $\pi_i = (n-1)(n-2)\ldots(n-i)/n^i$, $i = 0, 1, \ldots, n-1$.
For large n we have $ns \simeq 2K$,

$$E(T - t_o) = \frac{2K}{n}\sum_{i=0}^{n-1}\pi_i \simeq K\sqrt{\frac{2\pi}{n}} \quad \textit{cf. } \text{Cane, p. 179}$$

$$\simeq \sqrt{\pi}Ks \quad \ldots \quad (6)$$

and

$$var(T) \simeq \frac{4K^2}{n}\left(2 - \frac{\pi}{2}\right)$$

$$\simeq 0{\cdot}86\, Ks. \quad \ldots \quad (7)$$

From (6) and (7) we see that T^* may be estimated by

$$\bar{T} + \frac{1{\cdot}16}{s}var(T) - 1{\cdot}91\sqrt{var(T)} \quad \ldots \quad (8)$$

where $var(T)$ is the observed variance.

Using formula (9), the threshold was estimated from the data presented by some investigators, and the results are shown in Table I.

Discussion

If we assume that $p(t)$ is given by (5) and the hypothesis of Hardy, Wolff and Goodell (1952) that T^* is relatively constant over the population, then in this model t_o and K may vary with the individual only in such a way that $t_o + K$ is constant. Therefore, in (9) $var(T)$ is the variance of the stopping points for a given subject, or the within subject variance if more than one subject is used. In those experiments where a subject has been used more than once the within subject variance has not been included in the results, so in the above analysis we have assumed that t_o and K do not vary within an experimental block, and we have used the block variance instead. It does seem that t_o and K vary between blocks.

The results of Chapman and Jones (1944) and Chapman, Cohen and Cobb (1946) have been included in Table 1 even though these investigators did not instruct their subjects as to the end-point of pain. The reason for including them is that the high thresholds of these subjects compare with those of 21 instructed subjects (Group C) tested by Haslam using a step-size of 8 mc./sec./cm.^{2} These four groups excepted, the estimates of T^* for the forehead area seem to be centred about 220 mc./sec./cm.^{2} The variance of the Group C results is large relative to other experiments in which a similar step-size has been used, and this raises the question as to whether the group is typical or not. Clausen and King (1950) report that T is large for some subjects because they do not experience a pricking pain at all but only a burning sensation following one of heat. Since the presence of such individuals in a small group of untrained subjects would tend to make the group unrepresentative, it may well be that the small size of Group C makes this set of data unreliable.

Under the hypothesis that T^* is constant, the standard deviation of the estimates of T^* (excluding the four groups mentioned above) is 15·8 mc./sec./cm.^{2} compared with 26·7 mc./sec./cm.^{2} for $\bar{T}$. So it can be argued that it has been worth making the correction on the basis of the above model, since doing so gives less variability than that of the estimates which originally formed the basis for belief in the hypothesis. However, the

TABLE I

PAIN THRESHOLDS IN THE (*a*) FOREHEAD AREA; (*b*) FOREARM AREA

Investigator	*No. of observations*	*s*	*s.d.*	$\overline{T}$	*T**
(*a*)					
Schumacher, Goodell, Hardy and Wolff (1940)	324	20† (10)	21·0	206	192 (217)
Clausen and King (1950) ..	31	10	22·8	183	200
	31	10	22·4	181	196
Birren, Casperson and Botwinick (1951) ..	50	10	30·2	184	232
	50	10	29·0	180	223
	50	10	30·6	187	238
	50	10	32·0	183	241
Hardy, Wolff and Goodell (1952)	63	20	27·0	235	226
Haslam (1965)	21	32	43·3	248	233
	21	32	23·1	228	203
	19	16	22·4	238	231
	16	16	20·6	230	221
	11	8	33·1	193	289
	10	8	41·7	166	339
Chapman and Jones (1944)	200	20†	45·0	305	337
Chapman, Cohen and Cobb (1946)	44	20†	24·0	287	275
(*b*)					
Clausen and King (1950) ..	31	10	18·6	193	198
	31	10	15·2	187	185
Hardy, Wolff and Goodell (1952)	63	20	32·0	210	209
	63	20	24·0	200	188

† Step-size used was not mentioned by investigator.

hypothesis needs to be more firmly established in view of the finding of Clausen and King that some difference may exist between the pain thresholds of different parts of the body (see Table I), and that these are not constant differences between individuals.

I wish to thank Miss V. R. Cane for drawing by attention to the problem and for very helpful discussions.

REFERENCES

BIRREN, J. E., CASPERSON, R. C., and BOTWINICK, J. (1951). Pain measurement by the radiant heat method. *J. exp. Psychol.*, **41**, 419–24.

BROWN, J., and CANE, V. R. (1959). An analysis of the limiting method. *Brit. J. statist. Psychol.*, **12**, 119–26.

CANE, V. R. (1956). Some statistical problems in experimental psychology. *J. roy. statist. Soc.*, *B* **18**, 177–201.

CHAPMAN, W. P., COHEN, M. E., and COBB, S. (1946). Measurements related to pain in neurocirculatory asthenia, anxiety neurosis, or effort syndrome. *J. clin. Invest.*, **25**, 890.

CHAPMAN, W. P., and JONES, C. M. (1944). Variations in cutaneous and visceral pain sensitivity in normal subjects. *J. clin. Invest.*, **23**, 81–91.

CLAUSEN, J., and KING, H. E. (1950). Determination of the pain threshold in untrained subjects. *J. Psychol.*, **30**, 299–306.

HARDY, J. D., WOLFF, H. G., and GOODELL, H. (1952). *Pain Sensations and Reactions.* Baltimore: Williams & Wilkins.

HASLAM, D. R. (1965). The influence of stimulus scale-interval upon the assessment of pain threshold. *Quart. J. exp. Psychol.*, **17**, 65–8.

SCHUMACHER, G. A., GOODELL, H., HARDY, J. D. and WOLFF, H. G. (1940). Uniformity of the pain threshold in man. *Science*, **92**, 110–2.

Manuscript received 26th October, 1965.

PAYOFF: A NEGLECTED FACTOR IN REACTION TIME MEASUREMENT

BY

JOHN ANNETT

From the Department of Psychology, Hull University

The aim of this paper is to point out some implications of implicit or explicit payoffs in reaction time experiments. In any psychophysical task the subject is given a set of instructions which define the desired performance, as it were the rules of the game. Instructions may be supplemented by knowledge of results which further define the required performance. When performance admits of being "good" or "bad," however wide these limits, the instructions may be said to define, explicitly or implicitly, a system of payoffs. In reaction time experiments the subject is instructed, amongst other things, to respond as soon as possible after the stimulus but never before it. These instructions could equally well be represented by a system of payoffs and Figure 1 illustrates the kind of payoff system normally implied in these instructions. If S represents the time at which the physical stimulus occurs the graph shows a fixed penalty for anticipations (response before S), maximum positive score for responses immediately following S declining at an arbitrary rate as a function of elapsed time between S and the response until an unacceptably long time involves a constant maximum penalty. Some experimenters choose to treat very short RT's as anticipations. This is represented by the dashed line in Figure 1, but it should be clear that payoff values are, to a certain extent, arbitrary.

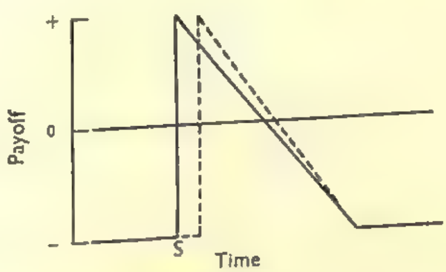


FIGURE 1

Payoff as a function of time between stimulus and response.

Reducing these instructions to a simple numerical score not only eliminates ambiguity and improves comparability between experiments but has the further advantage of permitting other factors to be combined with the instructions in a simple way. The customary catch trial rate, say 10 per cent., can be combined with a system of payoffs and we can then specify a single number representing the *average utility* of any response at a given time following, or preceding, S. This is simply the product of the score at time R and the probability of S.

The same procedure of combining instructions in the form of payoffs into a single number, the average utility of responses at time R, can be applied to other cases of stimulus uncertainty, for instance the temporal uncertainty when S is preceded by a warning signal at a variable interval. Figure 2 shows a warning signal S_w at time O followed by a signal to respond at varying intervals distributed in a rectangular fashion (i.e. equiprobable) between S_1 and S_n. The average utility of a response at any time R is obtained simply by summing the products of signal probability at times S_1 to S_n and the payoff for each possible S–R or R–S interval. A *utility function* can then be computed showing the average net score for all relevant values of R. Each set of conditions, catch trial rate, warning interval distribution and a specified payoff system can be combined to give a utility function unique to that set of conditions. Figure 2 shows three typical utility functions superimposed on a rectangular warning interval distribution. It is

assumed that S can occur at any one of 10 equally spaced times between S_1 and S_n and that the scoring goes from $+3$ for RT < 100 millisec. to -3 for RT > 699 millisec. with a penalty of -3 for anticipations. Curve A shows the average utility of responses where the range of warning intervals is 1 sec. and the catch trial rate is 10 per cent. Curve B shows utility for a range of 0·5 sec. and a catch trial rate of 10 per cent. and curve C shows utility for a warning interval range of 1 sec. and a catch trial rate of 50 per cent.

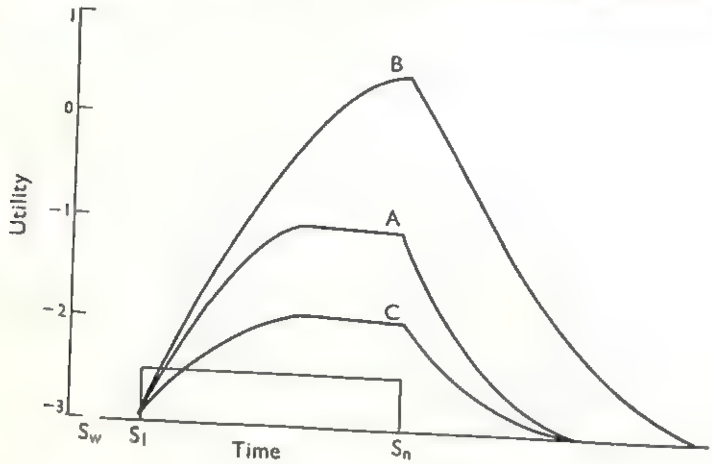


FIGURE 2

The utility of responses as a function of warning interval distribution, payoff and catch trial rate.

A	Warning interval range 1 sec.	Catch trial rate 10 per cent.
B	,, ,, ,, 0·5 sec.	,, ,, ,, 10 per cent.
C	,, ,, ,, 1 sec.	,, ,, ,, 50 per cent.

These curves are merely illustrative but in general it can be said that given a rectangular warning interval distribution and the type of payoff illustrated in Figure 1 *the utility of responses is invariably lowest at the beginning of the warning interval range and reaches a maximum towards the end of the range*. Although stimuli are equiprobable over the range *responses are not equally profitable*. Only when the range of intervals used is large compared with the maximum acceptable RT is the utility curve flat for any appreciable part of its length and this only holds when the penalty for anticipation is equal to the penalty for unacceptably short RT. The initial rise in the curve can only be offset by an appropriate increase in the frequency of signals at the shortest warning intervals. The catch trial rate affects the shape of the curve little, depressing utility overall, and it would appear that with specified payoffs the catch trial is not a particularly valuable form of experimental control.

This approach has interesting implications for the "refractory period" experiment and the variant in which the first stimulus is simply a warning signal. Figure 2 in fact represents the conditions of such an experiment. If the subject interprets his instructions as indicating a system of payoffs similar to that shown in Figure 1 he will find that it is rather risky to respond near to the beginning of the warning interval range and he will, on average, do better when responding towards the end of the range. Here the risk of anticipation is small and there will be a fair proportion of fast RT's to compensate for the occasional slow responses. The subject wishing to maximize payoff should ensure that his responses occur towards the end of the range and *it will then necessarily follow that the shortest warning intervals give the longest RT's* and that faster RT's and most of the anticipations will be found to occur at the longest warning intervals. This is typically the result of these experiments. Although in any given case we cannot be sure of the exact form of the utility function since payoffs have never been numerically specified, the usual conditions will give curves of the form illustrated in Figure 2. Thus it may be that the phenomenon of refractoriness is not due to any blocking of the central channel but simply to the subject following instructions.

Manuscript received 8th January, 1966.

THE EFFECT OF GRIP-TENSION ON TACTILE-KINAESTHETIC JUDGEMENT OF WIDTH

BY

G. BURN EVANS and EDGAR HOWARTH

From the Department of Psychology, University of Alberta

The effect of various grip-tensions on the accuracy of kinaesthetic width judgements was tested. Forty first-year psychology students were used as subjects. Significant differences in accuracy between pressures were found in the descending adjustments with greatest accuracy at 1·0 kg. Some significant differences were found between pressures in ascending adjustments. All subjects overestimated on the descending and underestimated on the ascending trials at all pressure levels. Increased grip-tension was found to reduce the accuracy of width judgement in terms of constant error while affecting variance only slightly.

Introduction

There have been relatively few studies of kinaesthetic width judgement and there appear to have been none on the particular parameter of tactual pressure, although Costello (1962) suggested the possibility that this parameter might proportionately affect felt width. Studies which have not controlled this parameter, but which presumably used light touch in the region less than 1 kg., have found DLs of the order of 1/20–1/50 using various psychophysical methods and widths ranging up to 100 mm. (Gaydos, 1958; Stevens and Stone, 1959; Dietze, 1961). Jastrow (1899) reported DLs as low as 1/100 in his best subjects.

In an initial experiment of this kind it might be unwise to attempt to make definite predictions because a number of possible outcomes could be envisaged. Some of these possibilities are listed below.

Where width is kept constant and pressure applied by the hand across that width is varied, how do the additional muscular and kinaesthetic sensations affect judgement of the felt width? One possibility is that the additional information could assist the judgement. A further possibility is that the additional sensations may hinder judgements in two ways: (*a*) by acting as distracting or competing stimuli, or (*b*) by serving to activate the stress mechanisms which might have a Yerkes–Dodson effect on the judgement.

Systematic errors, as between ascending and descending series, might occur in which subjects would tend to overestimate the standard width on descending trials, and underestimate on the ascending (Costello, 1961).

Finally, it occurred to the authors that if a light touch (approximately 1 kg.) was applied across the adjustable comparator and if different tensions were applied across the 3 in. standard, then the judgements might be most accurate when the tension across the comparator was approximately equal to that applied across the standard width. Therefore, because of the scarcity of previous relevant studies, it was decided to proceed on an empirical basis and leave the data to confirm the appropriate hypothesis.

Method

Apparatus

A more complete description has been provided elsewhere (Howarth, 1964) and only general features will be given. The apparatus consisted of two pieces; (1) The standard stimulus object consisted of two parallel brass plates, 3 in. apart, and 8 in. long, by 2 in. deep. A set of springs was placed between the plates, to provide different pressure levels as required (1, 2, 3, or 4 kg.). To ensure that subjects maintained the appropriate pressure, a contact-operated buzzer was attached to the standard so that, if the pressure were released, the plates were forced apart by the spring, breaking the contacts and causing the buzzer to sound. (2) The adjustable comparator was similar in length to the standard but had parallel sides which subjects could adjust to any width between 2 in. and 4 in. by operating a foot pedal which controlled a motor in the comparator. The motor moved the plates closer or further apart, depending on which side of the foot pedal was pressed. Pressure between the sides of the comparator was always less than 1·2 kg., as greater

pressure prevented the motor from operating, and subjects were instructed to use a light touch on the comparator.

The apparatus was placed so that the standard was on the right and the comparator on the left of each subject, with the sides of the two pieces parallel to, and on either side of his median plane, at a height just above knee level. The inner faces of the standard and comparator were 36 in. apart.

Subjects

Forty psychology students, 25 men and 15 women, from an introductory psychology class, served as subjects. They all attended for 1 hr. as part of the course requirement.

Procedure

Each subject was blindfolded and instructed to hold the comparator lightly between the forefinger and thumb of the left hand, and the standard similarly with the right hand. The procedure involved two stages.

Stage 1 consisted of 16 estimations of the widths of two blocks of wood ($2\frac{1}{2}$ in., $3\frac{1}{2}$ in.). The purpose of this phase was to make a preliminary estimate of each subject's accuracy of width judgement, and all those making absolute errors greater than 10 per cent. were eliminated from the experiment. Four subjects failed to meet this criterion.

Stage 2. In the second stage, subjects were required to make 24 judgements of the width of the standard, using the method of adjustment, with three settings at each of the four pressure levels, for both the ascending and descending series. In each of the three replications (four pressures × ascending and descending series × three replications) the conditions of pressure and series were randomized for each subject. For an ascending adjustment, the comparator was pre-set by the experimenter at approximately $2\frac{1}{2}$ in., for a descending adjustment at $3\frac{1}{2}$ in. Each adjustment by the subject took approximately 45 sec. He removed his hands from the apparatus between adjustments, and this was a signal to the experimenter that each judgement was completed.

Only three of the subjects were left-handed, and since there are conflicting reports in the literature (Costello, 1961; Wertheimer, 1954) on this problem, this data was included. Left-right randomization was not undertaken because of the nature of the apparatus and its electrical connections.

Results

The three readings taken at each pressure were averaged for each subject, taking the ascending and descending series separately. Table I shows means and standard deviations for the matchings of the 3-in. block under various pressures.

TABLE I

Matchings of 3-in. Block under Various Pressures

	Descending series				Ascending series			
Pressure (kg.) ..	1	2	3	4	1	2	3	4
Mean width (in.) ..	3·11	3·15	3·22	3·24	2·93	2·92	2·91	2·91
S.D.	0·11	0·14	0·15	0·15	0·13	0·15	0·20	0·20

For clarity of presentation we shall present the ascending and descending series separately.

(1) *Descending series*. Constant error ranged from 0·11 in. to 0·24 in. Trend analysis showed a significant linear component ($F = 44$, *d.f.* 1, 117, $p < 0{\cdot}01$) with a non-significant quadratic component. The direction of the trend showed increasing constant error with increasing tension. Variability of matchings showed little change with changes in pressure. Duncan's Multiple Range Test (Edwards, 1962) was applied to the mean estimations, and it was found that all the means differed significantly from each other with a protection level of 93 per cent. Analysis of variance applied to the descending series data supported the findings of Duncan's test.

(2) *Ascending series.* Constant errors ranged from 0·07 in. to 0·09 in. Trend analysis showed non-significant linear and quadratic components. Significant differences in variability were found between the matchings made for 1 kg. and 3 kg. ($F = 2{\cdot}24$, *d.f.* 39, 39) and between 1 kg. and 4 kg. ($F = 2{\cdot}25$, *d.f.* 39, 39). Duncan's test showed significant differences between these latter means. Analysis of variance showed no overall significance between the means.

Discussion

Several possible outcomes were discussed initially and it must now be asked which of these is favoured by the data. It appears that a somewhat different picture emerges according to the direction of adjustment. In the descending series the effect of increasing pressure was to increase constant error in the judgements without affecting the variance of the settings, whereas, in the ascending series, pressure did not affect constant error but to some degree influenced the variability of the settings. Because of these differences between ascending and descending series, an unequivocal statement, favouring one prediction over the other could not be made, although it does appear in general, that the smaller pressures resulted in more accurate judgements, expressing these in terms of either constant error or variability.

However, this finding should be treated with caution because of the possibility that width matching is easier when the pressures are equal on each hand. In the present experiment this occurred with a pressure of 1 kg. across the standard (3 in. block) when matched by a comparator pressure of less than 1·2 kg. One way to investigate this problem further would be to see whether acccuracy of width matching was affected by pressure when the width to be matched was a visual width, and cross-modality tests in both directions (visual to kinaesthetic and vice versa) are currently being undertaken.

Differences in the pattern of judgement between ascending and descending series agree with those reported by Costello (1961) wherein subjects overestimated block width in descending series and underestimated in the ascending series.

The general size of the DLs obtained agreed with those of previous workers. Some reported DLs are: Jastrow (1899) 1/100; Gaydos (1958) 1/25–1/50; Dietze (1958) 1/20–1/25; Stevens and Stone (1959) 1/20–1/30. This study found DLs ranging from 1/15 to 1/30.

This study was supported by National Research Council (Canada) Grant No. APA-85. The apparatus was made in the University of Alberta Physics workshop by Mr. N. Riebeck. Our thanks are due to Prof. W. N. Runquist for his comments on the original draft, and to Mr. P. DeGroot for construction of ancillary apparatus.

References

Costello, C. G. (1961). Constant errors in measurement of kinaesthetic figural after-effects. *Amer. J. Psychol.*, **74**, 473–4.

Costello, C. G. (1962). The effects of meprobamate on kinaesthetic after-effect. *Brit. J. Psychol.*, **53**, 17–26.

Dietze, A. G. (1961). Kinaesthetic discrimination: the difference limen for finger span. *J. Psychol.*, **51**, 165–8.

Edwards, A. L. (1962). *Experimental Design in Psychological Research.* New York: Holt, Rinehart and Winston.

Gaydos, H. F. (1958). Sensitivity in the judgment of size by finger span. *Amer. J. Psychol.*, **71**, 557–62.

Howarth, E. (1964). An apparatus for measuring kinaesthetic width judgments. *Amer. J. Psychol.*, **77**, 492.

Jastrow, J. (1899). *Mind.* (Cited by Dietze).

Stevens, S. S., and Stone, J. (1959). Finger span: ratio scale, category scale, and JND scale. *J. exp. Psychol.*, **57**, 91–5.

Wertheimer, M. (1954). Constant errors in the measurement of figural after effects. *Amer. J. Psychol.*, **67**, 543–6.

Manuscript received 25th January, 1966.

APPARATUS

AN E.C.G. GALVANOMETER USED AS A SHUTTER

BY

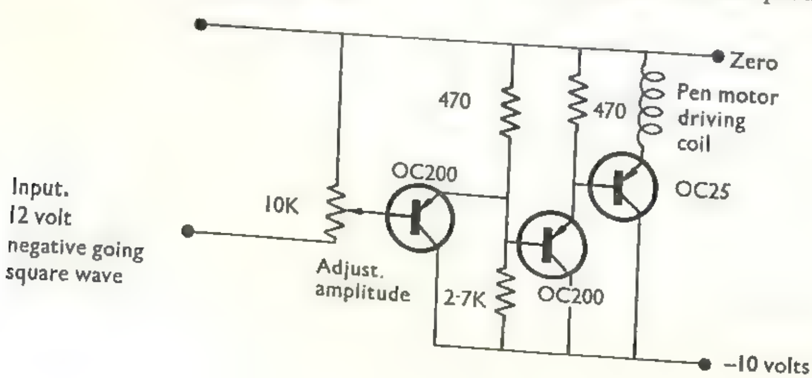
H. B. MORTON* and M. E. WILSON†

*From the *Department of Applied Electrophysiology, National Hospitals, Queen Square, London, W.C.1, and †Neurophysiology-Biophysics Research Unit, Boston VA Hospital, Boston, Massachusetts, U.S.A.*

Experimental psychologists frequently require a means of forming a high intensity light stimulus whose duration can be precisely varied from a few millisec. up to several sec. Unfortunately light sources which can themselves be switched rapidly on and off, are either of low intensity (e.g. cold cathode fluorescent tubes with fast decay phosphors) or low total light output (e.g. glow modulator tubes) or else they produce flashes which last only a few microsec. (e.g. electronic flash tubes). Light flashes must therefore be produced by placing a shutter in the path of light from a continuously lighted high intensity light source. It is desirable that such a shutter should be mechanically simple, robust, silent in operation and easy to install in a straight line optical system. It should also produce light pulses having brief rise and fall times and durations readily varied over a wide range of values. Most of the shutters generally available, fail to meet one or more of these requirements (see below) but the device to be described has all the above characteristics.

The shutter consists of a small piece of aluminium foil mounted on the writing arm of an Electrocardiograph Pen Motor (a moving coil galvanometer). The galvanometer is activated by the application of square electrical pulses from a galvanometer amplifier circuit which is itself driven by a pulse generator.

The angular velocity achieved by this type of shutter is limited by the weight of the writing arm as well as by the characteristics of the galvanometer. The authors found, however, that a piece of aluminium foil 1 cm square, mounted 8 cm. from the centre of rotation of a writing arm which had been lightened by the removal of the ceramic tip and the internal wiring, would achieve a mean angular velocity of 50 radians/sec. for at least 0·5 millisec. The edge of this shutter was therefore able to traverse a light beam 2 mm. wide, in 0·5 millisec. The rise and fall times of flashes of light so produced were equal in duration (0·5 millisec.) and independent of the total exposure time, provided this exceeded 5 millisec. A shutter used by one of the authors was operated in excess of 50,000 times in the course of 20 months without noticeable alteration in the shape of the light pulses. Light pulse durations were very constant over periods of several weeks. This shutter is therefore suitable for the production, over long periods of time, of light pulses varying in duration from 5 millisec. to more than 1 sec. and having a total rise and fall time which never exceeds 1 millisec. The exposure durations are readily and accurately controlled by the pulse generator. The shutter can also conveniently be used either to produce trains of pulses of any desired mark-space ratio by the application of appropriate sequences of electrical pulses to the motor, or to "gate" trains of light pulses produced more conventionally by a rotating sector disc.



The driving coil of the pen motor used by the authors* had a resistance of 7 ohms and an angular sensitivity of 2·4 amp. per radian, necessitating currents of the order of 0·2 amp. for the displacement described above. This current can be supplied through a transistor, or a relay switched by a suitable timing circuit. The galvanometer amplifier shown in the figure is operated by 12 volt negative-going pulses, but other suitable combinations of pulse generator and amplifier could be employed, or simple timing circuits built for particular applications. Oscillation of the shutter at the extremes of movement may be damped either mechanically by suitable stops, or electrically by providing a low driving source impedance and if electrically damped, the shutter is silent in operation. This shutter is readily available and relatively inexpensive (the pen motor costs £30). It is robust, simple to install and easy to regulate.

The convenient combination of characteristics provided by this shutter is not fully matched by any other device available.

Polarizing shutters (e.g. Kerr cells) are suitable only for microsecond pulses, and further, the ratio of light transmission in the open to that in the closed position is only in the order of 100 : 1. The exposure durations produced by conventional camera shutters are continuously variable, and in consequence exposure times may vary considerably during repeated resetting. In the authors' experience the characteristics of these shutters do not remain constant during prolonged use, and they are greatly affected by the rises of temperature often encountered in optical apparatus.

Rotating sectors, while ideal for producing trains of identical light pulses, can be made to give single flashes only after considerable elaboration. It is cumbersome to alter the duration of the pulses of light which are produced either by rotating sectors, or by falling blind or mirror shutters, especially if this alteration has to be carried out repeatedly or if the ratio of durations is large.

The performance of the Electrocardiograph Pen Motor shutter is most closely matched by other electromechanical devices. Relays may be made to power a shutter by attaching to their armatures, extensions carrying opaque squares. If one relay exposes the light beam, the second occluding it again, a simple shutter having a large amplitude of movement may be constructed, the duration of the light flash depending on the temporal interval between the electrical pulses activating the two relays. Unfortunately, however, this conveniently large amplitude of movement is obtained only at the expense of adding the extra weight of an extension to the armature, thus increasing the rise and fall times of the light flash. The times taken for the shutter blades to expose or occlude a light beam 2 mm. wide were found by the authors to be in the order of 5 millisec. Furthermore, chance differences in the position in which the armatures came to rest after activation, altered the time interval between the application of the following electrical pulse to the relay, and the consequent movement of the armature. This caused variation in the duration of successive light pulses: a variation which was sometimes sufficient to reduce the duration of a nominally 7 millisec. flash to zero.

Finally, electromechanical vibrators and loudspeaker movements, while having a frequency response which is initially superior to that of the Electrocardiograph Pen Motor when used as a shutter, provide a very restricted range of movement: in the order of ±0·05 in. If so small a movement is to be directly coupled to a shutter then the shutter must be very precisely aligned in the optical system at a point at which the image of the light source is very small indeed (no more than about 1 mm. in diameter). The latter requirement is often inconvenient especially in a Maxwellian view system, since it entails subsequent remagnification of the image of the light source. Mechanical magnification of the movement of the vibrator requires great technical skill if the weight of the levers is not to greatly impair the frequency response of the shutter. Alternatively, the movement of the vibrator may be optically magnified by mounting a mirror on its moving part. Activation of the vibrator may then displace, through an aperture, a light beam reflected from the mirror. Such a device, however, is not readily incorporated into a straight line optical system.

An Electrocardiograph Pen Motor and the associated timing equipment may, with minimal modification of its writing arm, be used to provide controlled movement of a projected stimulus, the transparencies being mounted on the arm. It has also served as a tactile stimulator and further uses can be envisaged.

Manuscript received 8th January, 1966.

* A standard E.C.G. pen motor made by Devices Ltd. A similar motor is made by, among others, Sanborn Co., in U.S.A.

OBITUARY

GEORGE HUMPHREY (1889–1966)

While not a member of the Experimental Psychology Society, George Humphrey was a well-wisher who often attended our meetings and occasionally reviewed in this *Journal*. He addressed the Experimental Psychology Group at a meeting held at Cambridge in April, 1951.

An Oxford man, Humphrey went to Leipzig to study experimental psychology just before the first War and later gained his Ph.D. at Harvard. In 1924, he was appointed to the Charlton Professorship of Philosophy at Queen's University, Kingston, Ontario, where he built up an active psychological laboratory in which much of his own best work was done. He returned to Oxford in 1947 as the first Professor of Psychology in the University and succeeded Dr. William Stephenson as Director of the Institute of Experimental Psychology in the following year. After his retirement, Humphrey lived for some years in Cambridge, where he was a frequent and welcome visitor to the Psychological Laboratory.

Although Humphrey's output of experimental work was not large, some of it is very well known indeed. In particular, his papers on habituation have been described as classical. Humphrey also worked on conditioning and reinforcement, on audiogenic seizures in the rat, and on apparent visual motion.

Of Humphrey's books, by far the best is *The Nature of Learning*, published in 1933. This remarkable book did much to set the study of learning in its biological context and had wide influence in its day. His later book on *Thinking: Its Experimental Psychology* (1951) never quite got to grips with the problems in the same way but has none the less proved useful as a source book and historical survey. After his retirement, Humphrey edited *Psychology through Experiment* and, with Michael Argyle, its companion volume *Social Psychology through Experiment*, both of which have been widely used as introductory texts.

Although he achieved distinction, George Humphrey never perhaps had quite the success he deserved. He was an unassuming man who hated intellectual assertiveness and perhaps for this reason did not always make the best of his opportunities. None the less, he will be widely remembered as an upright, shrewd and lovable man.

O. L. Zangwill.

BOOK REVIEWS

Annual Review of Psychology. Vol. 17, 1966. Edited by Paul R. Farnsworth, Olga McNemar and Quinn McNemar. Palo Alto, California: Annual Reviews Inc. Pp. x + 589. $9.00.

The 1966 volume of the *Annual Review* contains 17 articles and includes two new titles: "Status of Japanese Experimental Psychology," by Yoshihisa Tanaka, and "Cognitive Functions," by J. P. Van de Geer and J. F. M. Jaspars. Two advertised chapters (Personality Dynamics; Electrophysiology and Behavior) do not appear.

Contributors to the *Annual Review* must have great difficulties in deciding what to leave out of their chapters: very few of the titles could be comprehensively treated in the available space. Showing their awareness of this, most of the authors announce an intention to restrict the scope of their review. In some instances, however, the proposed limitations seem less than severe, as in Kagan and Henker's review of developmental psychology where "Primary emphasis was given to work with human or animal subjects that had a developmental implication" (p. 1). Other writers have a more radical solution, and both Ward (Audition) and Poulton (Engineering Psychology) simply describe work they believe to be important and interesting. This measure is entirely successful in both cases, as it is in Hernández-Peón and Sterman's "Brain Functions." Faced with an impossible title these authors have elected to concentrate on their own research interests (Attention and Sleep) and the chapter thus contains useful detail and genuinely critical comment. Another almost meaningless title is "Cognitive Functions"; Van de Geer and Jaspars are forced to write nearly four pages defining their intended subject, only to end with the decision "to leave out anything that is traditionally covered in other chapters of the *Annual Review of Psychology*" (p. 149).

On the whole, the best solution to the problem of selection would be to leave out uninspiring work and to concentrate only on studies that tell us something new. This step is not always taken. It is often difficult to tell whether the work is being reported accurately, so uninformative does it seem. The following, for instance, is hardly a novel finding: "Sherif and Sherif observed the group behavior of adolescent boys in natural settings for a period of five to seven months. The boys' conscious concerns were with cars, girls, and sex; power was a critical dimension in group structure" (p. 31). Even less exciting is the description of a "genuine innovation" in psychotherapy that involved the use of "operant conditioning methods to get the adolescent to the interviews" (p. 71). In essence, the trick was to give the subjects pay, cigarettes and food if they attended: youngsters, in other words can be bribed. A salesman would agree that this was good "Psychology," but would be puzzled to find it described as a genuine innovation. "Operant Conditioning," by the way, is becoming a very popular phrase; it might help readers to know that it means giving people a reward if they do the right thing, and bears little relation to carefully controlled and analytical studies using the same name.

It is perhaps a little naughty to poke fun at this sort of work: there is a danger of calling results trivial that would have seemed equally obvious had they gone the other way. This can be said in defence of many of the experiments reviewed by McGuire (Attitudes and Opinions), for instance the finding "that political slogans made more of an impact on subjects eating a free lunch than on subjects smelling unpleasant odors" (p. 482). Even so, it is difficult to find significance in the observation that "foremen and executives spend a great proportion of their time interacting with other people rather than working alone" (Porter: Personnel Management, p. 401). Foremen, after all, are paid to interact with other people, and this is only another example of "Operant Conditioning."

The lack of solid achievement in some of the subjects is honestly described by the reviewers themselves. Porter is particularly pessimistic about Industrial Psychology, and notes that "In some senses, at least, we always seem to be about 10 years behind the time" (p. 415). Other writers are more hopeful: Yates (Psychological Deficit) believes that a new approach is near to finding the nature of the defect in schizophrenia. This approach is to "conceptualize the human operator as an information-processing unit" (p. 127). The subsequent discussion describes ideas that are not nearly as vague as this introduction would suggest. Progress is also claimed for Behavioral Genetics (McClearn and Meredith), though it is not immediately apparent that we should be pleased by the "impressive advances in metabolic errors in mental defect" (p. 516).

I have not been able to mention all the contributions; they include helpful accounts of Color Vision (De Valois and Abramov), Psycholinguistics (Ervin-Tripp and Slobin) and Somesthetic Senses (Sherrick). This volume of the *Annual Review* is quite as useful as its predecessors, and the only general criticism is that the writers often scatter their efforts over too wide a range of topics. It is true that many authors have the sense to ignore the rather vague chapter headings, but a changed editorial policy is needed, if only to protect us from contributors who are tempted to take their terms of reference seriously.

M. J. Morgan.

Pavlov's Typology. Compiled, edited and translated by J. A. Gray. International Series of Monographs in Experimental Psychology, Vol. 1. Oxford and London: Pergamon Press. 1964. Pp. xviii + 480. 84s.

In Western countries the writings of I. P. Pavlov have had rather little influence on subsequent research. In the U.S.S.R. on the contrary they have been immensely influential, and in neurophysiology it can hardly be doubted that their influence has been mainly bad. It has deflected investigators away from the search for substantially new phenomena and into the pursuit of trivial details in a field, the classical conditioned reflex, which Pavlov himself had by 1917 taken about as far as its interest deserved.

The relative neglect in the West of Pavlov's work on conditioned reflexes does not rest merely on ignorance; Anrep's translation of 1926 made the main facts easily accessible to English-speaking readers. But the research on behaviour-types in dogs which occupied much of Pavlov's energy in the last 15 years of his life has until now been largely concealed in a language difficult and forbidding even to those Western readers who can read the Russian of the great novelists and of modern physicists and chemists with ease and pleasure. J. A. Gray's book does much to end this concealment. He has translated into astonishingly good English (considering the muddiness of the originals) an account by B. M. Teplov of Pavlov's work on behaviour-types, and five recent papers from

Teplov's laboratory. He has added two chapters of his own. The first reviews the Russian work (mainly from Teplov's laboratory) which has examined differences in behaviour between people, compared them with differences between dogs, and tried to interpret them in Pavlovian language, a language whose most striking feature is that it uses, to refer to input-output relations, words that seem to imply particular intervening mechanisms, but never specify these mechanisms clearly enough to allow the statements made about them to be treated as hypotheses and sharply tested. This defect of Pavlovian language and Pavlovian thought is one of which Gray is well aware, and his second chapter is an attempt to substitute definite hypotheses expressed in modern neurophysiological terms for those hinted at, but never clearly stated, by the Russians. My attitude to this attempt is coloured by my belief that the task was not worth attempting; but if it was worth attempting, then Gray has done it well. He has not completely bridged the immense gulf between Pavlovian language and that of Western neurophysiologists, for his reinterpretations of Pavlovian concepts are not yet precise enough to be tested experimentally, but they make a big step in the right direction.

G. S. Brindley.

Cognitive Processes and the Brain. Edited by Peter Milner and Stephen Glickman. Insight Books, No. 30. Princeton and New Jersey· Van Nostrand. 1965. Pp. viii + 215. 15s. 6d. ($1.95)

This latest "Insight book" contains 13 papers, some slightly abbreviated, from the periodical literature, all but three of which were published in the last six years. Although most are concerned with animal experiment, a few deal with neurological deficits in man, among them Hebb's admirable and somewhat inaccessible review of the effects of early brain injury on intellectual development. The papers are prefaced by short summaries composed by the Editors, which are for the most part helpful, though it is rather shaming to learn that psychologists have apparently to be told the meanings of terms such as sclerosis or oedema. Although the Editors have undoubtedly chosen well, one may regret that students should be thus spoonfed, or should one say trained by the automatic delivery of intellectual food pellets?

O. L. Zangwill.

Functions of the Corpus Callosum. Edited by E. G. Ettlinger. Ciba Foundation Study Group, No. 20. London: Churchill. 1965. Pp. xii + 156. 20s.

In the absence of Dr. Roger Sperry, this one-day Symposium held in London last year must have been a little like a performance of *Hamlet* without the Prince of Denmark. None the less, Dr. George Ettlinger did extremely well in gathering together so many first-rate workers and in organizing this admirable discussion. Apart from studies directly employing the split-brain technique, mention should be made of Professor Whitteridge's stimulating observations on inter-hemispheric visual organisation and Professor Jeeves' interesting studies of human patients with presumed agonesis of the corpus callosum. The discussion is uniformly high-level.

O. L. Zangwill.

An Index of Responses to the Group Rorschach Test. Studies of the Psychological Characteristics of Medical Students II. By C. B. Thomas, D. C. Ross and E. S. Freed. Baltimore: Johns Hopkins Press. London: Oxford University Press. 1966. Pp. xiv + 502. £6.

This atlas-sized volume, like its predecessor published in 1964, stems from Rorschach tests given to Johns Hopkins medical students as part of a long-term study of precursors of hypertension and coronary disease begun 20 years ago. It appears to be a labour of love, and one may sincerely hope that love's labour will not be entirely lost.

It is interesting to speculate on what Hermann Rorschach might himself have thought of the astonishing edifice which has been built on the foundation of his modest and tentative psychodiagnostic studies. Although he regarded his work as merely a beginning, can he really have anticipated the fantastic lengths to which it has since been taken? Confronted with this immense volume, one may feel that he would have had difficulty in restraining a wry smile.

O. L. Zangwill.

Perception and the External World. A reader edited and with an introduction by R. J. Hirst. London and New York: Macmillan. 1965. Pp. viii + 310. 15s. ($1.95).

This book comprises a short but lucid account of the kernel of the problem of perception by Professor Hirst. Each major point he makes is then elaborated by a reprinted essay by some authority: these essays comprise the bulk of the book.

Hirst describes first what the problems are, and emphasizes that many philosophers today do not deal adequately with the psycho-physiological accounts of perception. Hence, in Part I he presents the problems as discussed by one philosopher and three scientists (Brain, Helmholtz, Koffka). In Part II he presents discussions of the main philosophical theory that has been put forward to deal with this scientific evidence, as well as to try to give a descriptive account of what is actually going on in perception. This is the "sense-datum" theory: the case for is put by Moore, Price, Russell and Ayer, and the case against by Austin and Ryle. In Part III he presents discussions of what he regards as the fundamental problem of perception "Whether it is concluded that the immediate data in perceptions are sense-data or percepts, the fundamental problem is: What is their relationship to external physical objects?" He presents here two theories: Locke on Representative Realism and Sellars on Critical Realism, together with comment on these by Russell and others. The last section deals with radical theories such as Berkeley's and Phenomenonalism. Hirst grants that the representative theory has certain clear-cut advantages, but he points to several well-known defects. (However, recent work on the logic of representation may have made these difficulties less formidable). Hirst states that critical realism avoids some of the difficulties of the representative theory, but that it has been unable to present a convincing and comprehensive theory of perception, because, he argues, this part of its programme "cannot be undertaken without a full investigation of the psychological processes in perception."

Finally, he concludes that Phenomenonalism remains for many quite unconvincing, so that today we are in an impasse. The Direct Realism, to which many philosophers have returned, cannot "counter the scientific evidence nor provide a satisfactory explanation of hallucinations" and yet, as we have seen, the other forms of Realism have their own particular difficulties. However, Professor Hirst is not downcast by this and sees this impasse as an intriguing and stimulating challenge to philosophical endeavour.

This book may be warmly recommended to readers of this *Journal* as it presents the position with a sophisticated understanding (which some philosophers tend to lack) of the problems raised by experimental psychology for philosophy, and should make psychologists take note of their own philosophical position relevant to perception—for we must all have one, whether this be manifest or latent.

J. R. SMYTHIES.

Mood and Personality. By Alden E. Wessman and David F. Ricks. New York: Holt, Rinehart & Winston. 1966. Pp. xi + 317. $5.95 (cloth); $4.50 (paper).

This is a new book, not a reprint. The work described was done at the Prince House Annex, Centre for Research in Personality, the successor to H. A. Murray's Harvard Psychological Clinic. Indeed it is very much in the tradition of Murray's *Explorations in Personality*, first published in 1938 and reprinted in paperback a few years ago.

Basically, the authors have been concerned to chart mood level and its fluctuations in undergraduates of both sexes over a 6 weeks period of study. This involved the construction of some new "personal feelings scales" on which the subjects rated themselves daily. An attempt was also made to see whether people's ideas of themselves (as appropriately specified) underwent significant changes with shifts of mood-level, and if so, whether these were in the directions predicted by certain psychiatric theories. Finally, the authors tried to discover whether there are any general personality characteristics that distinguish people who are on the whole happy from those who are on the whole unhappy, or people whose moods undergo little short-term fluctuation from those who show greater variability. In long ago days this inquiry would have been called a study of temperament.

The methods used comprised several of the better known personality tests, both psychometric and projective, and a good deal of reliance is placed on the results of factorial analysis. Considerable use is also made of the Q-sort technique which we owe to William Stephenson—who, incidentally, had a good deal to do with the inspiration of this project. But the authors are not averse to using biographical and personal interview data to fill out the picture.

The most important finding seems to be that general level of mood (on the elation-depression dimension) and short-term fluctuation of mood appear to be relatively independent of one another. It thus makes sense to ask (*a*) what conditions appear to govern the general level of mood; and (*b*) what conditions appear to govern short-term mood fluctuations. With regard to (*a*), the authors claim that their findings indicate that people who regard themselves as for the most part happy tend to be optimistic, self-confident, controlled and decisive, whereas people who regard themselves as unhappy

tend to be pessimistic, low in self-esteem and dissatisfied in their personal relationships.

With regard to (*b*), evidence is presented that people little subject to mood variation tend to be independently-minded, reserved, consistent in their standards, and to show some tendency to repress or deny feeling. Those subject to appreciable mood variation, on the other hand, are apt to suffer more from inner difficulties but to show notably greater responsiveness and originality.

In their interpretation of these differences, the authors make much play with the concept of "ego-strength" and its relation to the quality and stability of mood. It is not clear, however, whether the latter are to be regarded as cause or effect of variations in "ego-strength." In general, the authors tend to stress constitutional and genetical rather than environmental factors but do not seem very clear as to how these are to be envisaged.

This study has considerable freshness and interest. One might regret, perhaps, that the problems were conceived quite so broadly and that no attempt was made to link individual differences in mood with constitutional factors of other kinds, e.g. physique. Thus it would not have been difficult to combine the psychological study with somatotyping and to see whether any of the differences observed could be correlated with the varieties of temperament adduced by Sheldon. In the case of the women subjects, too, closer attention might well have been paid to premenstrual alterations in mood level.

Although the authors were clearly anxious to restrict their inquiry to the study of normal people, certain of the problems in which they are interested might have been better studied in a psychiatric setting. For example, changes in self-assessment correlated with mood shifts are much more apparent in patients in the course of recovery from affective illness than in the normal. A technique which allowed these to be properly assessed might be of considerable value in psychiatry.

It should be said that anyone who may be tempted to use the authors' "personal feeling scale" with undergraduates in this country should take warning that some of the items may not be unproductive of ribaldry. For example, the present reviewer must admit to some hesitation in assessing his own more encouraging moods with regard to criteria such as "replete with life's abundant goodness" or "I am a surging torrent of spectacular insights." Somehow, the language of elation is altogether too florid to be taken quite seriously.

Although a serious attempt to study the dimensions of temperament is to be greatly welcomed, the present reviewer is not quite convinced that this particular inquiry is altogether on the right lines. He would have preferred to have seen more use made of ratings by others as well as self-ratings, and to have been given more information about physique, family history and other possible indications of constitutional status. He would also have welcomed closer analysis of the relations between mood and actual personal experiences during the 6-week periods of study. None the less, this is one of the very few recent studies devoted to the exploration of temperament and may well provide a groundwork for the more systematic study of affective disorder. O. L. ZANGWILL.

Personal and Organizational Change Through Group Methods: The Laboratory Approach. By Edgar H. Schein and Warren G. Bennis. New York and London: Wiley. 1965. Pp. x + 376. 62s.

The laboratory methods of inducing social change which are described in this book are the product of field theory methods applied to encapsulated social situations—social engineering married to scientific awareness. The results of the match are fascinating and have already affected 50,000 people according to the authors' claim. This is a conservative measure of the method's influence, since the people concerned have mostly been of considerable influence in a wide variety of organizations.

This, then, is an important area of human psychological endeavour which Schein and Bennis have tried to encompass, and the result is an interesting mixture of sophisticated social science and "movement"-oriented material. The most common type of laboratory is a residential 2-week period in which the entire social experience of the participants is carefully planned to maximize learning. There are many ingenious variations on this basic method which are well described.

Evaluating the changes induced by the method are identical to those present in other forms of education and therapy, and the techniques which are presented here show once more just how difficult this can be. In this respect the book is likely to fail in its aim of

acquainting a wider audience with the method for which simpler accounts are available, but for the specialist the reports are very valuable.

The enthusiasm of the contributors shines through most of the essays, but it is a great pity that they diminish this with a pseudo-language. After all, who needs to add "deroutinization" or "learnings" to their vocabulary? This is not carping criticism, since the use of such terms mostly serves to provide stones for the many existing opponents to throw.

DOUGLAS HOOPER.

Laboratory Experience in Psychology: A First Term's Work. By B. Babington Smith. Oxford: Pergamon Press. 1965. Pp. xxiv + 240. 25s.

In the late 1920's Mr. Babington Smith was introduced to "practical psychology" at the University of Cambridge. He suffered then, as he says, "a deep frustration" and "saw repeatedly the disruption of the good will and interest which beginners bring to the laboratory." In 1932 he joined the psychology department at St. Andrews University, was given charge of practical teaching and resolved to try to develop a course that would inspire and retain the interest of his students and lead them to a genuine understanding of the science of psychology. He continued in this resolve until he removed to the University of Oxford in 1947, and subsequently for a further 10 years. This book describes in great detail, and with abundant illustration, the introductory course which his reflexion led him to develop. The story is one of wide interest and could be studied with advantage by anybody who has to try to introduce students to experimental work in psychology.

Babington Smith claims that his design for a first term's experiments was dominated by three principles: (1) the student must always understand what he is aiming at, or at least must "have an intelligible aim"; (2) he must proceed "from the familiar to the unfamiliar"; (3) each step in succession must "appear to follow intelligibly from positions already reached."

His own classes began with a series of Demonstrations arranged to illustrate "The Evidence of Our Senses." A student reports his own evidence in detail, and the collected reports are discussed by supervisor and students together. His selected Demonstrations cover matters such as the after-effects of colour; appropriateness of names to visually presented forms; effects and after-effects of watching rotating spirals; some of Michotte's well-known displays of the apparent relations of forms in movement. Any instructor could, however, make a different choice according to his own preferences and available equipment.

Following the Demonstrations a set of experiments are carried out; there is a final recapitulatory session. How to arrange and discuss the experiments are exhaustively described. Together the exercises complete the First Term's Work. The leading idea is for experiment to begin on a descriptive level and gradually to become more and more abstract, "controlled" and quantitative. Thus the topics considered in this book are (1) Serial Reproduction; (2) studies in the control of Meaning, including normal forgetting, association, nonsense syllable control, recognition; (3) "getting beneath recognition," mainly through tactile experiences; (4) the perception of weight; (5) Extent and Density and their parts in perceptual learning; (6) Length in relation to "frames of reference." As with the Demonstrations, any supervisor could make a different choice of specific experiments, provided he adopted the same kind of progression from qualitative to quantitative, and from descriptive to statistical. Babington Smith's account of what he and his students did, how he endeavoured to direct and help in the treatment of results obtained, and his own observations of the significance of the work for psychological understanding is written out in very great detail. It could be condensed with advantage at a good many points, and it is perhaps a bit of a pity that no attempt is made to show by actual evidence just how the course proposed produces effects different from and better than the more traditional approaches. There is however no other book that makes as serious and thoughtful an attempt to deal with the problems involved, and the volume deserves serious consideration from all who are concerned in the teaching of experimental psychology.

There is a lively—and very controversial—"Editor's foreword" by Professor Patrick Meredith.

F. C. BARTLETT.

Animal Drives. Edited by George A. Cicala. London and Princeton: van Nostrand. Insight Book No. 31. 1965. Pp. vi + 265. 15s. 6d. ($1.95).

This is another collection of journal reprints, being aimed this time at the physiological

psychologist. Seventeen papers grouped under three headings deal with the physiology of drive (five papers), its causes (seven), and its effects (five). Only two of the reports (including one by the anthologist himself) originally appeared within the last 10 years, and it is hard to see how a collection like this could possibly live up to the publishers' claim to cover the "present state and direction" of the subject. The subject has of course, been advancing at an almost explosive rate during the past few years.

Although the volume comprises an otherwise fairly predictable selection of names such as Mayer, Brobeck, Olds, Sheffield, Skinner, Hull and N. E. Miller, there is scant attention to theoretical work but a rather puzzling emphasis (four papers reprinted in their entirety) demonstrating "the belief that hunger drive increases with increasing food deprivation"—hardly the most exciting discovery in the field. Most of the reprints are from A.P.A. journals, and all but three would probably be available in the average departmental library.

L. J. HERBERG.

Vision in Vertebrates. By Katharine Tansley. London: Chapman and Hall. Science Paperbacks. 1965. Pp. vii + 132. 12s. 6d.

The title of this book is slightly misleading, for vertebrate vision encompasses much more than an author can deal with in a slim paperback. It is therefore important to state the kind of material discussed by the author, and her approach to the subject. A great deal of anatomy and physiology is included, almost all of it concerned with the eye. For example, there are chapters on the anatomy and histology of the eye, retinal physiology, the visual pigments, and accommodation. The brain is hardly mentioned. The structure and physiology of the eye are throughout related to the behaviour of the animal, but it is behaviour chiefly in the sense of Natural History and Ethology. For instance, the visual apparatus is related to the style of life of the animal, such as nocturnal or diurnal habit, hunter or predator, aquatic or terrestrial. There is little reference to psychological experiments on vision, for whether a rat can discriminate between various patterns, whether monkeys without a visual cortex can see, why man experiences visual illusions, are all questions outside the scope of the book.

Nevertheless, this book should attract a fairly wide audience, chiefly introductory students of physiology and zoology, but also psychologists who would like to know about the eyes and habits of animals that are rarely studied in a psychological laboratory. Many difficult and controversial topics are mentioned rather than discussed, for example, the opponent processes theory of colour vision. But whenever this occurs reference is made to original papers in what is an excellent bibliography. Indeed, it is refreshing to see so few secondary sources cited in the references. The student is even referred to Young's 1807 book for an account of his trichromatic theory. This may deter the layman, but is just what the serious student needs.

There is a brief and final chapter on "What do animals really see?" Its message is that a knowledge of the anatomy and physiology of the eye is not sufficient to describe or account for visual behaviour, which must be settled by behavioural experiments. This nicely restores the balance.

ALAN COWEY.

The Psychology of Childhood and Adolescence. By C. I. Sandström. London: Methuen. 1966. Pp. 259. 36s.

Professor Sandström is best known to British psychologists as an editor of the *Scandinavian Journal of Psychology* and of *Acta Psychologica*, but the present book is addressed to a wider audience. In its original Swedish edition, it enjoyed a wide success both among teachers in training and the general public, and it can therefore be judged only in terms of the audience at which it is aimed. It is at too low a level for undergraduate courses in developmental psychology, while students of education and others needing to acquire a systematic knowledge of child development can turn to somewhat more rigorous but still readable textbooks—Lovell's *Educational Psychology and Children* being a case in point. As a Methuen *Manual of Modern Psychology*, Professor Sandström's book is a disappointment. Incredibly, it contains no references whatsoever, except a score of "suggestions for further reading" at the end of the book.

The orientation is broadly biological, and there are some useful discussions of imprinting and the work of the ethologists, though to say that their work "provides a comprehensive empirical complement to the psychoanalytical approach" (p. 88), may strike some as far-fetched. Professor Sandstrom has, unfortunately, chosen to follow Gesell in giving a year by year account of child development, which leads him to say of

the four year old, "The child changes rapidly from shyness to frankness, which makes him socially unstable" (p. 51). He follows an outmoded orthodoxy in subscribing to a rigidly genetic viewpoint involving I.Q. constancy ("A person with . . . the average quotient of 100 in his fourth year should keep this average at later testings" (p. 137)).

Other inaccuracies include the statement that deaf babies do not babble, and the use of M. K. Smith's 1926 norms on vocabulary growth which have been frequently shown to be inaccurate, and are largely superseded by Templin's work; moreover, sex differences in language development are no longer taken for granted. Spelt's work on "foetal conditioning" is enthusiastically quoted as an example of early learning, though the study was poorly controlled and rejected by authorities such as Peiper.

Despite some serious deficiencies, the book is a painless introduction to the subject, even though it cannot be recommended to the serious student. The photographs are delightful.

P. MITTLER.

Drives, Affects, Behavior. Volume 2. Edited by Max Schur. New York: International Universities Press; London: Bailey Bros. & Swinfen, 1965. Pp. 502. $10.00. (90s.)

This is a volume of essays in honour of the late Marie Bonaparte. While the contributions are almost entirely on psycho-analytical topics, experimental psychologists might be interested in a paper by Dr. G. S. Klein on hearing one's own voice and an account of psychological testing of Tibetans and Todas by H.R.H. Prince Peter of Greece and Denmark.

O. L. ZANGWILL.

Nonparametric Trend Analysis. By George Andrew Ferguson. Montreal: McGill University Press. 1965. Pp. 61. $1.95.

An experimenter often wants to know whether his data show an increase or a decrease over a range of conditions, or whether the relation is U-shaped or has some other form. This little book is a practical guide to nonparametric techniques for such "trend analysis." The statistic "S" is used, which is perhaps most familiar as a constituent of "Kendall's tau," but is in fact of very general application. As the author points out, the Mann-Whitney and Wilcoxon tests can be seen as special cases of techniques described here.

The book is compact, clear and untechnical. Two criticisms can be made. Firstly, a newcomer to trend analysis would benefit from more discussion of the pitfalls in the interpretation of tests for U-shaped or zig-zag trend. Secondly, the problems arising from tied data could well have a fuller discussion. There is no explicit consideration of this for the small-sample case, and there are objections to the procedure proposed for large samples. (See, for example, Jonckheere in *Biometrika*, 1954.) Nevertheless, this is a useful supplement to the standard elementary texts on nonparametric statistics.

PAUL WHITTLE.

Elementary Applied Statistics: for Students in Behavioral Science. By Linton C. Freeman. London and New York: Wiley. 1965. Pp. x + 298. 53s.

This is an introductory textbook, requiring no mathematical background and written on the assumption that formulae frighten people. As a consequence, it will be too gentle and wordy for some readers. Both parametric and nonparametric methods are given. The layout is unusual, being based on the treatment of all statistical questions as problems of association. Thus testing the significance of a difference between two groups becomes a special case of testing strength of association between two variables, one nominal (the dichotomy) and one ordinal (or whatever is appropriate). It seems doubtful that this arrangement has sufficient advantages to justify its use in an introductory textbook, but it may appeal to those who like to regard science from Professor Freeman's viewpoint, as "the study of relationships among variables."

PAUL WHITTLE.

La Psychologie Expérimentale. By Paul Fraisse. Paris: Presses Universitaires de France. 1966. Pp. 128.

Despite the paperback boom, nothing quite like the French "Que sais-je?" series exists in England—a collection of over one thousand monographs covering an immense range of topics, each written by a distinguished expert giving a survey of his particular field. This is popularization at its best for a price of a few shillings a volume. Professor Fraisse, in one of the latest additions to the series, succeeds in outlining the type of problems which confront experimental psychologists and the methods they use to tackle them, in a readable but far from superficial manner, by taking a selection of experiments

on specific problems. The book should do much to remove misconceptions about the role of the psychologist in the minds of the French reading public, and it is perhaps a pity that there are not more good popularizations of this kind in English.

R. DAVIS.

Deranged Memory. By G. A. Talland. New York and London: Academic Press. 1965. Pp. 356. 76s.

In this excellent book, which summarizes 6 years of hard work and the contents of several previous publications, Professor Talland describes the differences between a carefully selected group of 29 Korsakoff patients and two groups of matched controls drawn from psychiatric and orthopaedic hospitals. The author's main aims were:

(1) to determine and compare the actual capacities and skills in the two groups, and
(2) to test various hypotheses about memory disorder.

Clinical descriptions of the orientation, mood, affect and confabulation typical of the Korsakoff state are followed by an excellent summary of the literature on the subject given in its historical context. The author's ability to read these papers in the original and to distil their contents deserves our deepest gratitude. The various theories which have been put forward to account for amnesic states are critically discussed.

It is in the experimental section, however, that readers of this *Journal* will clearly be most interested. The intelligence, judgement and reasoning as well as the linguistic, drawing, visuo-motor and perceptual skills of the Korsakoff patients were compared with the control groups on a large number of tests. No significant differences were found as long as all the data necessary to solve the task were present at a time the solution was required, i.e. if retention and recall of data were not involved.

It was, as anticipated, on the memory tests that the groups showed complete divergence. Whatever type of test they were given the Korsakoff patients seemed to take in less than the controls and to forget faster. These are not in themselves new findings but Professor Talland's discussion of their implications is refreshingly original. He points out that to call the Korsakoff syndrome a disturbance of memory is merely to beg the question. What is Memory? The accepted phases of registration, retention and recall are not clearly distinguished in the amnesic state; for in these patients information that appears to have been forgotten will often emerge in unexpected forms or places. It is by studying the actual behaviour of Korsakoff patients rather than by trying to list their defects that one may hope to gain a clue as to what has gone wrong, and if their difficulties can be explained by any one single causation, Professor Talland concludes that it would be by the loss of (or premature closing of) a hypothetical "general activating function."

MOYRA WILLIAMS.

Scripta Hierosolymitana. Publications of the Hebrew University, Jerusalem. Vol. XIV. Studies in Psychology. Edited on behalf of the Department of Psychology by Rivka R. Eifermann. London: Oxford University Press (Jerusalem: Magnes Press). 1966. Pp. 230. 43s.

This volume is divided into four parts, dealing respectively with child development, social behaviour and personality, tests and measurement, and cognitive processes. Only the last two are likely to be of direct interest to readers of this *Journal*, in particular Eifermann's article on selection strategies in concept attainment and Schlesinger's on stimulus redundancy in a card-sorting task. Other papers are concerned with the development of smiling in infants from diverse environments (e.g. family, institution and Kibbutz), the growth of motor control in relation to intelligence, and a study of reactions in Israel to the Eichmann trial. Although the breadth of psychological interest revealed by these studies is encouraging, it is little disappointing that experimental psychology as understood today seems hardly to have got off the ground. O. L. ZANGWILL.

The Psychology of Ego-Involvements: Social Attitudes and Identifications. By Muzafer Sherif and Hadley Cantril. London and New York: Wiley. Science Editions. 1966. Pp. viii + 525. 19s. ($2.45).

This is a paperback reissue of a book which created quite a stir when it was first published 19 years ago. Its impact was probably due to the authors' attempt to bring greater realism into social psychology while retaining the advantages of quantitative treatment. In spite of the fact that the lessons it taught have been largely learned, it still remains bright, fresh and readable. O. L. ZANGWILL.

REVUE CANADIENNE CANADIAN JOURNAL DE PSYCHOLOGIE OF PSYCHOLOGY

VOLUME 20, NO. 2 JUNE/JUIN 1966

Published for the Canadian Psychological Association by
UNIVERSITY OF TORONTO PRESS

THE CANADIAN JOURNAL OF PSYCHOLOGY is published quarterly. Annual subscriptions, $6.00, single number $1.75.

Contributions. Original manuscripts and correspondence on editorial matters should be sent to:

The Editor, Canadian Journal of Psychology, McMaster University, Hamilton, Ontario.

Subscriptions to Journal. Orders and correspondence regarding subscriptions, changes of address, and purchase of back numbers shall be sent to:

Canadian Psychological Association,
The Secretary-Treasurer,
Suite 210, 225 Lisgar Street,
Ottawa 4, Ontario.

THE QUARTERLY JOURNAL OF
EXPERIMENTAL PSYCHOLOGY

Vol. XVIII 1966 Part 4

PARTIAL REINFORCEMENT AND BREADTH OF LEARNING

BY

N. S. SUTHERLAND

From the Laboratory of Experimental Psychology, Sussex University

Two experiments on partial reinforcement were undertaken to test predictions made by a two process model of discrimination learning. In the first experiment rats were trained on a discrimination involving two relevant cues: one group (C) was trained on a 100 : 0 schedule, the other (P) on a 50 : 0 schedule. Both groups were then given transfer tests with the two cues presented individually; finally all animals were extinguished on the original training stimuli and on the single cue stimuli. During extinction there was a negative correlation between the number of correct responses made by individual subjects of Group C to each single cue; whereas the correlation was positive for subjects of Group P. The second experiment employed basically the same design, but subjects were trained with seven relevant cues. The results of transfer tests showed that subjects of Group P learned to attach the correct response to many more cues than subjects of Group C. This suggests that the breadth of learning is greater under partial than under consistent reinforcement. The results were predicted by the model of discrimination learning under test.

Recent work indicates that discrimination learning involves two distinct processes—learning to which dimension of the stimuli to attend and learning to attach the correct responses to the relevant dimension (Mackintosh, 1965; Sutherland, 1964*a*; Zeaman and House, 1963). Sutherland (1964*b*) has suggested that the effects of partial reinforcement on extinction may be explicable in terms of such a theory. It is assumed that animals have a number of different analysers each of which analyses a particular dimension of the stimulus situation. Only when an analyser is switched in can the animal learn to attach the correct responses to its outputs. It is further assumed that analysers are strengthened when their outputs can successfully predict reward and punishment or any other events of differential importance to the animal. Thus if a rat is learning to jump to black and not to white in a jumping stand, once the correct response attachments start to be formed it will consistently be rewarded after the output corresponding to black occurs from the brightness analyser and it will consistently fail to be rewarded after the output corresponding to white has occurred. The idea that analysers are only strengthened when their outputs are correlated with differential events of importance to the animal may explain why a secondary reinforcer is more potent when its presence has signalled reward and its absence non-reward than when animals have been given only rewarded trials in its presence (Notterman, 1951). A similar proposal has been put forward by Egger and Miller (1962) who suggest that only stimuli providing information to the animal become secondary reinforcers. The present hypothesis is plausible on more general grounds since from a teleological standpoint the most useful stimulus dimensions are those which can be used to predict events of importance to the organism.

In a consistent reinforcement (CR) situation, there will normally be many different stimulus dimensions present and whichever one the animal attends to, it can consistently secure reward by attaching the correct responses to the appropriate points on that dimension. During partial reinforcement (PR), however, no feature of the stimulus situation can ever be fully correlated with reward. We might expect that under CR the response will be attached only to the outputs from one or two dominant analysers since it has already been shown that in CR the more that an animal learns about one cue, the less it learns about a second (Sutherland and Mackintosh, 1964; Sutherland and Holgate, 1966). In extinction, animals must either extinguish the response to the few analysers used in training, or extinguish the analysers themselves. In PR training, however, since no analyser consistently predicts reward the animal should try out many different analysers: the response will come to be attached to many analysers, and in extinction the response must be extinguished separately to all those to which it has been attached. If we assume further that when responses are attached to many different analysers of low strength extinction takes longer than when responses are attached only to one or two analysers of high strength, then we have a new explanation for the effect of partial reinforcement on extinction. The present explanation is not necessarily inconsistent with others that have been put forward, since the partial reinforcement effect may well be overdetermined and be due to the operation of several different factors. Putting the above argument anthropomorphically, it is as though under PR animals notice more different features of the stimulus situation than under CR.

The present article reports two experiments undertaken to test this explanation of PRE: a more indirect test has been reported elsewhere (Sutherland, Mackintosh and Wolfe, 1965).

Experiment I

It has been shown that if rats are trained with CR to discriminate between two conjoint cues, and are later tested on performance with the two cues presented singly, there is a negative correlation between individual animals' performance on each cue (Sutherland and Mackintosh, 1964; Sutherland and Holgate, 1966): in other words the more an animal learns during training about one cue, the less it learns about the other. If, however, subjects are trained with each cue presented singly on alternate trials, then there is a positive correlation between the number of trials taken to learn to make the correct response to each cue. According to the hypothesis under test, animals given PR should not strengthen one analyser at the expense of all others: they should keep trying out different analysers in an effort to find one which yields consistent predictions about reinforcement. It should follow that subjects trained with two conjoint cues under PR should show a positive correlation between how much they learn about each cue: they will behave like subjects given alternating trials with two cues presented singly. The present experiment tests this prediction.

Apparatus

A modified Lashley jumping stand was used. Subjects were trained to jump from a Y-shaped jumping stand on to a ledge protruding in front of the stimulus windows in order to obtain food from a food well behind the windows. The apparatus was painted matt grey. Fuller details of the apparatus and experimental procedures appear in Sutherland and Holgate (1966).

Procedure

All subjects were kept on a 22·5 hr. feeding schedule, being given free access to food for 1·5 hr. each day 10 min. after they had been run. All subjects were pretrained to

jump to plain grey windows and push them down to obtain food. During training animals received 10 trials a day with an 8 min. intertrial interval. On rewarded trials subjects were given 15 sec. access to small pellets of Purina chow: water was also present in the food well. On trials on which no reward was given, rats found the stimulus window locked, and were removed from the ledge to which they had jumped after 5 sec.; no correction of a wrong choice was allowed. After training, transfer tests were given in alternation with retraining trials. On transfer trials, both windows were unlocked and animals always received a food reward whichever way they jumped. The order in which transfer tests were given was balanced so that each member of each pair of transfer stimuli appeared equally often on the left and on the right, and equally often on the same and the opposite side to the positive stimulus of the preceding retraining trial. Finally, all subjects were extinguished by running trials with both stimulus doors locked.

Subjects

The subjects were 19 male hooded rats about 4 months old at the start of the experiment.

Experimental design

The experiment was run in three stages.

Training. Subjects were trained to jump to a white horizontal rectangle and to avoid a black vertical rectangle. Thus they could solve the problem in terms of either the brightness or the orientation differences between the stimuli or in terms of both. They received 10 trials a day for 20 days.

Transfer tests. Subjects were next given retraining trials with the original training stimuli, alternating with transfer test trials on which one or other of the two single cues was presented. Forty transfer tests were given, 10 on each of the following pairs of rectangles: white horizontal vs. black horizontal, white vertical vs. black vertical, white horizontal vs. white vertical, black horizontal vs. black vertical. The first two pairs of transfer shapes test how much animals learn at the first stage about the brightness cue; the second two pairs how much they learn about the orientation cue. All rectangles measured 10 × 2 cm. and were made from $\frac{1}{4}$ in. Perspex glued to the centre of the grey stimulus windows.

Extinction. All animals were given 60 extinction trials (12 a day) with the stimulus windows locked. Twenty trials were given with the original training shapes and 10 with each of the four transfer pairs set out above. Every third trial was with the original shapes.

Ten subjects (Group C) were trained with CR (100 per cent. reinforcement for jumps to the positive stimulus, 0 per cent. reinforcement for jumps to the negative stimulus); nine subjects (Group P) were trained under PR and were given 50 per cent. reinforcement for jumps to the positive, 0 per cent. for jumps to the negative. The sequence of reinforcements and non-reinforcements for correct jumps made by Group P was determined by Gellermann orders. During transfer tests, all animals received the same reinforcement schedule on the retraining trials as they had received during original training, and they were always reinforced whichever stimulus they chose on transfer trials. During extinction, no reinforcement was given and rats were removed from the jumping platform if they failed to jump within 2 min.

Results

Training

Group C reached a criterion of all 10 trials correct within a single day in an average of 7·8 days, Group P in 13·7 days. The difference is significant by a Mann-Whitney U test at better than the 0·01 level. Over the last 40 trials of training Group C made 98 per cent. correct jumps, Group P 91 per cent.

Transfer tests

Group C scored 97 per cent. correct on retraining trials, 83 per cent. correct on black–white transfer tests, and 88 per cent. correct on horizontal–vertical transfer

tests. The corresponding scores made by Group P were 82, 67 and 68 per cent. If we combine the probability of a correct response occurring with each of the single cues to predict performance with two cues according to the formula

$$P = p_1 . p_2/((p_1 . p_2 + (1-p_1).(1-p_2))$$

where P is the probability of a correct response with both cues present, and p_1 and p_2 the probability of a correct response with the single cues present alone, then the retraining results are accurately predicted from the single cue transfer tests. The formula yields a prediction of 97 per cent. correct on retraining trials for Group C and 81 per cent. correct for Group P.

Product moment correlation coefficients were calculated between individual subjects' performance on the brightness and on the orientation transfer tests. For Group C, $r = 0.0$ and for Group P, $r = +0.33$. Although the difference is in the predicted direction it is not significant: moreover, the correlation for Group C was not negative as it was found to be in other experiments (Sutherland and Holgate, 1966). The reason for the failure to obtain a negative correlation for Group C may be that subjects had received so much overtraining on the original two-cue shapes that they had learned about both cues. Thus three animals of Group C made no mistakes on either cue.

Extinction

The results are summarized in Table I. Group P made significantly more responses than Group C: this is the normal PRE effect, and is explained on the present theory by the assumption that the response of jumping is conditioned to the outputs from more analysers in Group P than in Group C. The ratio of correct to wrong jumps is significantly higher both on retraining and on transfer trials for Group C than for Group P ($p < 0.01$ by Mann-Whitney U tests). This also is to be expected, since Group P should have the response of jumping conditioned to irrelevant cues more strongly than Group C. Thus, if Group P subjects adopt a position habit during training, they would be rewarded on 25 per cent. of trials as against 50 per cent. if they respond correctly to either of the two relevant cues. Since no cue enables them to predict reinforcement with complete consistency, the position analyser should be strengthened relatively more for Group P than for Group C: Group C can achieve 100 per cent. success if they attend to the relevant cues and only 50 per cent. if they attend to position. (The argument assumes that there is a bigger subjective difference between 100 and 50 per cent. success than between 50 and 25 per cent. success.) If, then, the response is strongly conditioned to irrelevant analysers in Group P, they will make a lower ratio of correct to incorrect

TABLE I

EXPERIMENT I—EXTINCTION

	Two cues			*White–Black*			*Horizontal–vertical*		
	R	*W*	*N*	*R*	*W*	*N*	*R*	*W*	*N*
Group P ..	67	8	25	57	18	24	58	13	29
Group C ..	64	1	35	46	3	51	46	8	46

R = Per cent. correct choices.
W = Per cent. errors.
N = Per cent. failures to jump.

responses than Group C since their responses will to some extent be determined by irrelevant analysers even when the relevant cues are present either together or singly.

If, in extinction, some responses occur when irrelevant analysers are switched in, we would not expect much difference between correlation coefficients for Groups C and P if the coefficients are calculated on ratios of correct to total responses made on the two sets of transfer tests. Such ratios will reflect not merely the relative amount learned about the two relevant cues but the strengths of irrelevant analysers and the strength of irrelevant analysers will affect ratio scores on both cues in the same way, thus tending to produce positive correlations. The correlations were in fact +0·65 for Group P and +0·21 for Group C. The difference is in the predicted direction but is not significant. Moreover, it is positive for Group C rather than negative. Once again no high correlation could be expected for Group C since there was very little variation in individual scores—three subjects in fact made no mistakes on either set of transfer stimuli. The *absolute* numbers of correct responses given to the transfer stimuli do vary considerably from animal to animal, and the strength with which each relevant analyser is switched in and the strengths of its response attachments will largely determine the absolute number of correct responses made to each single cue. Calculating r for absolute number of correct responses made by individual subjects to black–white and to horizontal–vertical, we obtain values of +0·80 for Group P and −0·86 for Group C. Both of these coefficients are significant at the 0·01 level of confidence, they are both in the predicted direction and the difference between them is significant at the 0·001 level. The prediction made by the theory is therefore strikingly confirmed.

Experiment II

A rather more direct test can be made of the theory than that reported in Experiment I. The most obvious prediction is that rats trained under PR will learn to attach the correct response to more features of the stimulus situation than rats trained under CR. We can discover whether this is true by training animals on a discrimination with many cues always present and relevant and then giving transfer tests with one cue at a time to find whether subjects trained under PR have indeed learned about more cues than subjects trained under CR.

Subjects and procedure

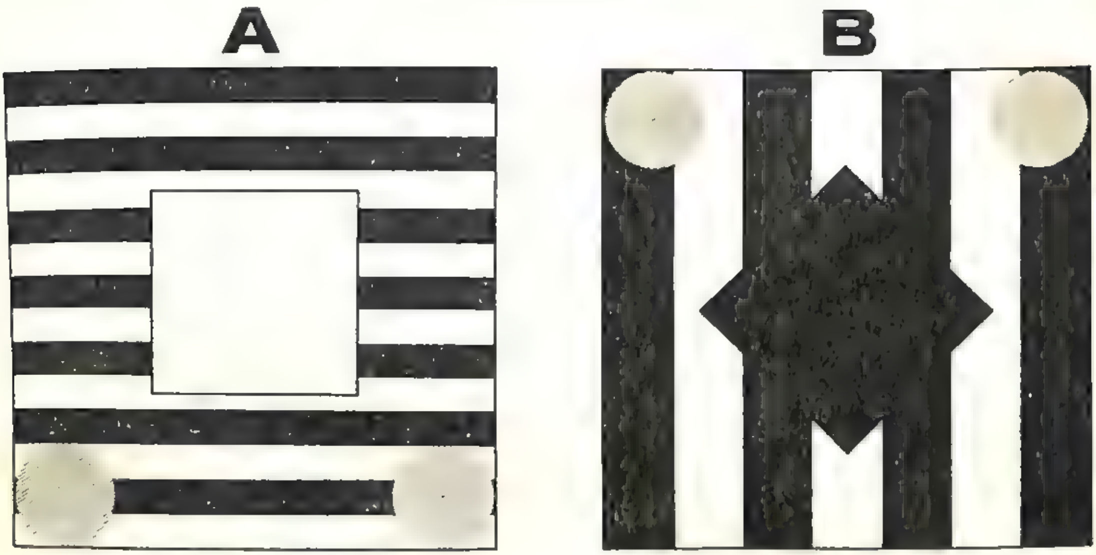
The subjects were 35 male hooded rats about 4 months old at the start of the experiment. The basic procedures used in training, testing, and extinction were the same as in Experiment I.

Training stimuli

The visual stimuli used during training are shown drawn to scale in Figure 1. The square, diamond and circles were made from ¼ in. thick Perspex and were respectively white, black and grey. The pieces of Perspex were mounted on top of the striped background. The area of the patterns was 14 cm.² and they covered the whole stimulus window. Thus there were five visual dimensions: horizontal vs. vertical stripes, narrow vs. broad stripes, white vs. black central shape, square vs. diamond, circles at bottom vs. circles at top. The visual dimensions were always combined in the way shown. In addition the arm of the Y-shaped jumping stand pointing to stimulus A was made of metal, and the arm pointing to stimulus B was made of rubber: the rubber and metal could be readily interchanged between trials and it was hoped that they would serve as a tactile cue, although rats may also have been able to see the difference between them. The seventh stimulus was provided by a bell which was sounded as each animal was placed on the jumping stand. The bell was behind and to one or other side of the jumping stand and all subjects had to learn to jump directly away from the bell—i.e.

to select the window on the opposite side of the apparatus to it. The bell was turned off 5 sec. after the subject jumped.

FIGURE 1



Training shapes.

Experimental design

Training. Eighteen rats were trained under PR, 17 under CR using the same schedules as in Experiment I. Half the subjects in Group P were trained with stimulus A positive, half with stimulus B positive; within Group C seven subjects were trained with A positive, eight with B positive. All animals were given 200 training trials, 10 a day, using a non-correction procedure. From trial 101 onwards, any subject that had jumped on all 10 trials of the previous day to the same side was given manual guidance to the non-preferred side when it was correct and reinforced until the position habit was broken (i.e. until it jumped once of its own accord to the non-preferred side).

Single-cue transfer tests. In the first series of transfer tests given immediately after the end of training, 20 trials were given with each cue present in isolation. When a visual cue was tested the bell was omitted and the floor of the jumping platform was made of wood on both sides. Thus horizontal and vertical stripes were shown with no other visual cue present: on 10 trials the stripes were broad, on 10 narrow. Similarly narrow vs. broad stripes were tested giving 10 trials with both stimuli having horizontal stripes and 10 with both stimuli having vertical stripes. Black vs. white was tested for 10 trials with two squares and for 10 trials with two diamonds of opposite brightness using a plain grey background on the doors. Square vs. diamond was tested with black shapes on half the trials and white on the remaining trials. The circles were presented at top and bottom of plain grey doors. To test metal vs. rubber and side of bell, plain grey stimulus doors were inserted. Ten trials were given a day and every other trial was a retraining trial with all seven cues present.

Double-cue transfer tests. The subjects and the experimenter were given a week's rest after the conclusion of the first series of transfer tests. After 2 days' retraining with the original shapes, animals were next tested with every possible pairing of two cues at a time: once again transfer tests were alternated with retraining trials. Twenty-one pairs of cues can be formed from seven single cues, and six transfer tests were given with each of the double cues. Again, 10 trials a day were given including retraining trials with the seven-cue stimuli.

Extinction. Finally, all animals were extinguished by giving 80 extinction trials to the training stimuli with both doors locked. Any subject that failed to jump within 60 sec. was removed from the apparatus, and extinction trials were discontinued for each subject after three failures to jump (not necessarily consecutive) within this time. At this stage only five trials a day were run.

Results

Training

Group C learned to a criterion of all 10 trials correct within a single day in a mean of 7·0 days; Group P required a mean of 12·5 days to reach the same criterion. Rats trained with shape A positive reached this criterion in 11·8 days and those trained with shape B positive in 8·0 days. An analysis of variance showed both effects to be significant, the first at the 0·001 level of confidence and the second at the 0·01 level; the interaction term was not significant. It is not known why subjects trained with stimulus B positive should have learned more rapidly than those trained with stimulus A positive. This effect did not show up at any of the subsequent stages of the experiment.

Eleven subjects from Group P and two from Group C required some manual forcing. Of the animals that received guided trials, those from Group C were given three each and those from Group P received a mean of 5·1 guided trials. There was no systematic relation between the number of trials needed to reach criterion and performance on transfer tests nor between whether or not a subject received guided trials and performance on transfer tests.

Over the last 20 trials of training, Group C made no mistakes; nine subjects of Group P also made no mistakes, six made one mistake and one made five mistakes. Thus all animals except one had mastered the problem thoroughly by the end of training.

Single-cue transfer tests

Although both groups as a whole performed best on horizontal–vertical, different subjects within each group performed best with different cues. Thus, if we consider the best score on any cue made by individual animals of Group P these scores are distributed over six of the seven cues; no animal scored best on the bell. The final two columns of Table II show the per cent. correct responses to cues selected according to individual subjects' performance: opposite cue 1 appears the mean of the highest scores made by each subject on any transfer test, cue 2 shows the mean of their second highest scores and cue 7 is the mean of each subject's worst score on any transfer test.

Considering first performance tabulated by actual cue present, it will be seen that although Group P performed worse than Group C on the retraining trials, they performed better on all transfer tests. All the group differences are significant at the 0·02 level of confidence or better, except for the difference in scores on horizontal–vertical which was not significant. Moreover, by a Mann-Whitney U test, Group P was significantly above chance with a confidence level of 0·05 or better on every single cue, whereas Group C was only significantly above chance only on the horizontal–vertical cue. There were no significant differences either on retraining trials or on transfer tests between the performance of subjects trained with stimulus A positive and those trained with stimulus B positive. Animals trained with shape A positive tended to do slightly better during retraining trials and with horizontal–vertical and metal–rubber than those trained with shape B positive, but neither these nor any other differences between groups trained in different directions were significant.

When the data are tabulated in terms of cues ranked according to individual subjects' performance, Group P again scored significantly better on every single cue except that on which performance was best. These results bear out the predictions from the theory with the exception that we might have expected that on

individual subjects' best cue Group P would do worse or at least no better than Group C since Group C should have switched-in the analyser for this cue very strongly as they are not learning about any other cue. In fact, Group P did slightly better than Group C on their best cue though the difference was not significant. This result might, however, be a statistical artefact: thus Group P averaged 66 per cent. on cues 2, 3, and 4, whereas Group C averaged only 54 per cent. on these cues. Since the best score for each subject on any cue is selected as cue 1, chance high scores on any of four cues might contribute to the score on this cue for Group P whereas because of the big drop between performance on the best cue and on the second best cue such scores will make a much smaller contribution to the cue selected as best for Group C.

TABLE II

Experiment II—Single-Cue Transfer Tests

Cue	*Group P*	*Group C*	*Cue*	*Group P**	*Group C**
Retraining trials	94	99			
Horizontal–vertical	71	67	Cue 1	76	70
Metal–rubber	69	54	Cue 2	67	59
Narrow–broad	64	51	Cue 3	66	54
White–black	59	48	Cue 4	64	51
Circles	59	47	Cue 5	59	49
Square–diamond	58	48	Cue 6	55	44
Bell	57	47	Cue 7	49	39

* These columns' scores are tabulated by taking the average of the highest scores on any single cue by each subject (cue 1), the average of the second highest scores (cue 2) etc.

It is evident from Table II that individual subjects from Group P have learned something about six cues. The subjects of Group C have only learned one cue to a level which differs significantly from chance. Thus, although they scored 59 per cent. on their second best cue, this could well be due to chance since in fact they scored 11 per cent. *below* chance on their worst cue: the below chance scores made by Group C on cues 6 and 7 are of course an artefact produced by the method of tabulating the scores.

The worse performance of Group C on the transfer tests was not due to their forming position habits more readily than Group P: indeed Group P actually showed more position habits than Group C when they were performing at chance or below. If we analyse by cues and by subjects, there were 27 occasions on which a subject of Group P scored at chance or below on a given cue, compared with 60 for Group C. On 67 per cent. of these occasions subjects of Group P showed a fixed position habit—i.e. on all 20 trials with a given cue they selected the same side; the comparable figure for Group C was 39 per cent. Thus the better performance of Group P on transfer tests occurred despite the fact that their performance on the original shapes was worse than that of Group C, and despite the fact that they were more prone to form position habits.

A product moment correlation coefficient was calculated between individual subjects' scores on horizontal–vertical and metal–rubber (the two cues which were best learned overall). The value of r was $+0{\cdot}61$ for Group P, and $-0{\cdot}11$ for Group C. Both coefficients are in the predicted direction and the first is significant at the 0·02 level of confidence. The difference between the two values is also in the predicted direction and is significant at the 0·02 level, thus confirming the results of Experiment I.

One further effect should be noted: there was a tendency for scores on transfer tests to decrease over trials. Thus Group P scored 76 per cent. correct with horizontal–vertical over the first 10 trials and only 67 per cent. correct over the last 10 trials. On the same cue Group C scored 73 and 62 per cent. over the same trial numbers. Group P showed some drop in scores between the first and last 10 transfer tests on all but their worst two cues; the shifts on the remaining cues were smaller for Group C and were variable in direction but they were in any case performing at near chance on all of them from the outset of training.

Double-cue transfer tests

Only six trials were given on each of the double-cue pairs of transfer tests because at this stage of the experiment animals tended to develop very strong position habits. The results are shown in Table III. The first two columns of figures were arrived at by calculating the average performance for each group over each of the six tests (36 trials) on which a given cue appeared. None of the differences between the two groups approaches significance. Many subjects, particularly in Group P, had developed very strong position habits. Thus out of 219 occasions on which a subject of Group P ran at chance or below on a given transfer test, animals ran to the same side on all six trials 78 per cent. of the time; out of 170 occasions on which subjects of Group C performed at chance or below, they selected the same side 55 per cent. of the time. The slight and non-significant tendency for Group C to perform better on transfer tests than Group P can be accounted for entirely in terms of the greater frequency of responses to position by Group P.

As in the single-cue transfer tests, both groups performed very much better on horizontal–vertical and metal–rubber than on any other cue. Table III also shows performance on the remaining cues after all trials on which horizontal–vertical appeared have been discarded, and after discarding both trials with horizontal–vertical present and with metal–rubber present. It will be seen that performance on the remaining cues is only slightly above chance. Group C did, however, perform better on all five of these cues than in their single cue transfer tests, and this is to be expected since although the results are analysed by single cues they are based on tests with two cues present. Group P, however, performed worse on each of the five cues when the results of tests involving horizontal–vertical and metal–rubber

TABLE III

EXPERIMENT II—DOUBLE-CUE TRANSFER TESTS

Cue	*All trials*		*Trials without horizontal–vertical*		*Trials without horizontal–vertical and metal–rubber*	
	Group P	*Group C*	*Group P*	*Group C*	*Group P*	*Group C*
Retraining trials	96	98				
Horizontal–vertical	75	80				
Metal–rubber	71	73	73	72		
Narrow–broad	62	62	61	64	55	63
White–black	58	60	62	60	56	54
Circle	58	58	58	58	52	51
Square–diamond	61	62	60	63	55	58
Bell	56	57	66	67	51	51

Figures indicate per cent. correct responses.

have been removed from the analysis. This finding is difficult to interpret since if animals have learned about seven cues conjointly there can be no reason why they should perform worse on the cues presented two at a time than when presented one at a time. The explanation of the poor performance of Group P at this stage must in fact lie in the development of very strong position habits.

Correlation coefficients were again calculated on individual scores on horizontal–vertical and on metal–rubber. The value of r was $+0{\cdot}20$ for Group P and $-0{\cdot}20$ for Group C. Both values are in the predicted direction, and the difference between the coefficients is also in the predicted direction but it is not significant.

Extinction

Some animals in each group were very slow to extinguish possibly because the bell was an aversive stimulus. Six subjects of Group P and four of Group C had failed to meet the criterion of extinction in 80 trials. If these are counted as having extinguished in 80 trials, then Group P averaged 67·3 trials to extinction, Group C 44·4 trials: a Mann-Whitney U test showed the difference to be significant at better than the 0·05 level of confidence.

Median latencies were calculated for each subject for each of the first 3 days of extinction. The median of the medians in sec. was Day 1—P 1·2, C 1·0; Day 2—P 1·2, C 3·0; Day 3—P 2·7, C 12·9. The differences between groups were significant on days 2 and 3 ($p < 0{\cdot}02$, and $p < 0{\cdot}01$ by a Mann-Whitney U test). Thus there was an effect of PR on extinction both as measured by latencies and as measured by trials to meet a criterion of extinction.

Discussion

Breadth of learning

The hypothesis to be tested in Experiment II was that rats trained under PR should attach the correct response to more cues than rats trained under CR. The results of single-cue transfer tests provide striking confirmation of this. The better performance of Group P on the less conspicuous cues is particularly remarkable since they performed worse than Group C on the original training shapes and they were more prone to form position habits than Group C.

There is a third factor which might be expected to militate against Group P performing better on transfer tests than Group C: Group C was effectively on the same reward schedule during retraining trials with the original shapes and during transfer tests, i.e. 100 per cent. reinforcement in transfer tests and almost 100 per cent. reinforcement on retraining trials. Group P, however, received slightly less than 50 per cent. reinforcement on retraining trials and 100 per cent. reinforcement on transfer tests. It might have been expected that the change of schedule would cause Group P to respond to the differences between transfer shapes and training shapes and therefore enable them to learn more readily that in transfer trials reward would be forthcoming no matter which way they jumped. This factor may have been responsible for the fall in performance of Group P during the first series of transfer tests and their subsequent very poor performance on double-cue transfer tests. One final difference between the groups should be noted. At the beginning of transfer tests, Group C had received at least twice as many reinforced trials as Group P. It is unlikely that the difference in performance during transfer tests was due to this factor, since Sutherland and Holgate (1966) have shown that increasing the number of training trials with two cues present leads to better performance when the less conspicuous cue is presented in transfer tests on its own.

Although none of the factors so far discussed could have produced the greater breadth of learning shown by Group P, there is yet a further possible explanation of the result which is more difficult to eliminate on the present data. It could be argued that if during training Group P subjects jumped with longer latencies than the subjects of Group C, they would have had more opportunity to notice multiple differences between the stimuli to be discriminated. Unfortunately latency data were not recorded except during the extinction phase of the experiment, and although at the beginning of extinction there was no significant difference in latencies between the two groups, this does not prove that there was no such difference during training. Two considerations render this explanation implausible: (i) Group P received only about half as many rewarded trials as Group C so that the effects of any differences in latencies would to some extent be counterbalanced by the greater number of trials on which Group C had the opportunity to learn. (ii) Group P developed stronger position habits in training and animals with position habits exhibit less VTE behaviour than animals without. Nevertheless a further experiment is necessary before this explanation can be completely eliminated.

Additivity of cues

The results for Group C present the following problem. It has been shown in Experment I and elsewhere (Sutherland and Holgate, 1966) that when subjects are trained with two cues present and then tested with one cue at a time, performance on retraining trials with two cues simultaneously present can be predicted by the following formula $P = p_1 \cdot p_2 / (p_1 \cdot p_2 + (1-p_1) \cdot (1-p_2))$ where P is the probability of making a correct response with both cues present and p_1 and p_2 are the probabilities of making a correct response with each cue present on its own. This formula (extended to cover the case of seven cues) and applied to the results of Group C on single-cue tests in Experiment II grossly underestimates the performance on retraining trials with all seven cues present. One possibility is that Group C had learned about compound cues: this could occur in two ways. They could either learn the compound positive and negative values of several cues and hence perform badly if the compounds were not present, or they could learn the positive and negative values of one cue but make the use of that cue dependent on the other cues being present regardless of their value. These two possibilities can be rephrased as follows, taking the first three stimuli as examples. The first possibility would mean that an animal learned to go towards the combination of horizontal, metal, and white and avoid the combination vertical, rubber and black. The second possibility would mean that an animal had learned that if rubber, metal, black and white are present it should jump towards horizontal and avoid vertical.

Some support for supposing that Group C had learned something about compound cues comes from the double-cue transfer tests. When two cues are simultaneously present, their scores improved markedly and to an extent that would not be predicted from the results with single cues alone. Thus although they performed at chance or below on the five least conspicuous cues presented singly, they scored above chance on four of them when they were presented in combination with the others taken two at a time. There is no evidence for whether or not Group P learned compounds in addition to components: the fall-off in their performance from the first to the second series of transfer tests is due to their developing very strong position habits and it is not possible to know what their performance on the double-cues would have been if these habits had not developed.

It should be noted that we cannot know for certain that the dimensions of the stimuli learned by subjects are exactly those that have been specified here. For

instance we cannot know whether animals learned in terms of stripe width or in terms of the number of stripes present, nor can we know whether rubber–metal was learned as a tactile or as a visual cue. This uncertainty does not affect the interpretation of the results, since, if we obtain above chance scores on transfer tests, the animals must have learned a dimension which is represented in the stimuli used in the transfer tests.

It seems reasonable to conclude that Group P had learned about more of the individual component cues than Group C, and to this extent the hypothesis under test is confirmed. It is unfortunate that the Group P results on double-cue shapes were vitiated by the development of position habits and the issue of how much is learned about compounds under different schedules of reinforcement can only be settled by further experiment.

Two further predictions from the two process theory of discrimination learning were confirmed. Because of the difference in reinforcement schedules, Group P did not learn about one cue at the expense of others; hence, in both experiments the correlations between how much was learned by individual subjects of Group P about different cues tended to be positive. The reason why positive correlations are obtained is presumably due to individual differences in general learning ability, health, motivation, visual acuity and so on. Because the more strongly one analyser is switched in the less strongly are others switched in, Group C tended to give negative correlations between the amount learned about different cues. Although the correlations in Experiment II were low, the important point is that there was a significant difference in the right direction between the values of r for the two groups.

For reasons given in the Results section of Experiment I, subjects trained with PR should be more prone to switch in irrelevant analysers than those trained under CR. This was again confirmed. In Experiment II, subjects of Group P were more prone to form position habits both in training and during transfer tests than subjects of Group C. The fact that in Experiment I during extinction, Group P made more incorrect responses than Group C and gave a higher ratio of incorrect to correct responses also suggests that they had attached the response of jumping during training to more irrelevant features of the situation.

Conclusions

In both experiments a significant effect of PR on extinction was demonstrated. It is this effect that the hypothesis was originally designed to explain, and it is important to demonstrate that in the situations used the PR effect on extinction does occur since if it did not, we would not expect to find any of the other phenomena predicted by the hypothesis.

It should be noted that the validity of a two-process model of discrimination learning does not depend upon its success in predicting results of partial reinforcement experiments. The main evidence for a two-process model comes from experiments on reversal learning and on controlling which analyser is switched in by other forms of pre-training (Sutherland, 1964*a*, Mackintosh, 1965). However, by assuming that the rules governing the strengthening of analysers are different from those governing the strengthening of response attachments, it is possible to understand PRE in terms of a two-process model. Although the rule suggested here for strengthening analysers is not sufficiently precise, the present experiments and the experiment reported in Sutherland, Mackintosh and Wolfe (1965) demonstrate that it is possible to apply the theory including this rule to partial reinforcement and to make new and interesting predictions.

This work forms part of a project on "Stimulus analysing mechanisms" supported by the American Office of Naval Research (Contract N62558–3927) and the Nuffield Foundation and undertaken at Oxford University.

The data analysis and writing up were undertaken whilst the author held a visiting appointment at M.I.T. This side of the work was supported in part by the Joint Services Electronics Program (Contract DA36–039–AMCO3200 (E)) by the U.S. Army Research Office, Durham; and in part by the National Science Foundation (Grant GP–2495), the National Institutes of Health (Grant MH–04737–05), and the National Aeronautics and Space Administration (Grant NsG–496).

References

EGGER, M. D., and MILLER, N. E. (1962). Secondary reinforcement in rats as a function of information value and reliability of the stimulus. *J. exp. Pshcyol.*, **64**, 97–104.

MACKINTOSH, N. J. (1965). Selective attention in animal discrimination learning. *Psychol. Bull.*, **64**, 124–50.

NOTTERMAN, J. M. (1951). A study of some relations among aperiodic reinforcement discrimination training and secondary reinforcement. *J. exp. Psychol.*, **41**, 161–9.

SUTHERLAND, N. S. (1964*a*). Visual discrimination in animals. *Brit. med. Bull.*, **20**, 54–9.

SUTHERLAND, N. S. (1964*b*). The learning of discrimination by animals. *Endeavour*, **23**, 140–52.

SUTHERLAND, N. S., and HOLGATE, V. (1966). Two-cue discrimination learning in rats. *J. comp. physiol. Psychol.*, **61**, 198–207.

SUTHERLAND, N. S., and MACKINTOSH, J. (1964). Non-additivity of cues. *Nature*, **201**, 529–30.

SUTHERLAND, N. S., MACKINTOSH, N. J., and WOLFE, J. B. (1965). Extinction as a function of the order of partial and consistent reinforcement. *J. exp. Psychol.* **69**, 56–9.

ZEAMAN, D., and HOUSE, B. J. (1963). The role of attention in retardate discrimination learning. In ELLIS, N. R. (Ed.) *Handbook of Mental Deficiency*, New York: McGraw-Hill. Pp. 159–223.

Manuscript received 31*st January*, 1966.

THE INFLUENCE OF ACOUSTIC AND SEMANTIC SIMILARITY ON LONG-TERM MEMORY FOR WORD SEQUENCES

BY

A. D. BADDELEY

From the Medical Research Council, Applied Psychology Research Unit, Cambridge

It has been shown that short-term memory (STM) for word sequences is grossly impaired when acoustically similar words are used, but is relatively unaffected by semantic similarity. This study tests the hypothesis that long-term memory (LTM) will be similarly affected. In Experiment I subjects attempted to learn one of four lists of 10 words. The lists comprised either acoustically or semantically similar words (A and C) or control words of equal frequency (B and D). Lists were learned for four trials, after which subjects spent 20 min. on a task involving immediate memory for digits. They were then asked to recall the word list. The acoustically similar list was learned relatively slowly, but unlike the other three lists showed no forgetting. Experiment II showed that this latter paradox can be explained by assuming the learning score to depend on both LTM and STM, whereas the subsequent retest depends only on LTM. Experiment III repeats Experiment I but attempts to minimize the effects of STM during learning by interposing a task to prevent rehearsal between the presentation and testing of the word sequences. Unlike STM, LTM proved to be impaired by semantic similarity but not by acoustic similarity. It is concluded that STM and LTM employ different coding systems.

Introduction

There has, until recently, been a tendency for short-term memory (STM), involving the retention of material over a period of seconds, to be studied using different techniques and concepts from those used to study long-term memory (LTM). The growth of interest in STM has recently changed this and produced the suggestion that LTM and STM lie on a single continuum and can be explained most economically in terms of one theoretical framework, that of interference theory (Keppel and Underwood, 1962; Melton, 1963; Postman, 1964).

Unfortunately, because of the difference in techniques traditionally used in the two areas, it is impossible to decide on the basis of the existing evidence just how similar LTM and STM really are. This decision can only be made when equivalent experiments are done in LTM and STM. One such attempt was that of Keppel and Underwood (1962), who used the technique developed by Peterson and Peterson (1959), in which a single verbal item is presented once followed first by an interpolated task designed to prevent rehearsal, and then by a retention test. Keppel and Underwood showed a decline in performance with successive stimuli, closely analogous to the phenomenon of pro-active inhibition (PI) in LTM (Underwood, 1949*a*). Subsequent studies have confirmed that an effect occurs, but suggest that it differs from the classical PI effect in a number of important respects. Loess (1964) confirmed that the first item of a series was best retained, but found no difference between the second and subsequent items, in direct contrast to the gradual build up of PI shown in LTM studies. A second prediction from LTM, that PI increases with length of interval between successive presentations was tested by Peterson and Gentile (1965), who found the opposite to be the case. Finally, a recent study by Conrad (1966) has shown that the nature of the intrusion errors shown in this situation is inconsistent with Keppel and Underwood's hypothesis namely that the forgetting shown is attributable to PI from previous items.

A further attempt to demonstrate classical LTM effects in STM was made by Baddeley and Dale (1966) in a study using paired-associate learning. Here a classical retro-active inhibition paridigm was used; in this the experimental group first learns A–B, then A′–C and is then retested on A–B, and traditionally shows poorer retention than the control group which learns A–B, followed by D–C before being retested on A–B. In the Baddeley and Dale study, A and A′ were adjectives with similar meanings and B, C and D were dissimilar adjectives. The LTM experiment showed the classical RI effect, more forgetting when the interpolated list was similar to the first list. In the equivalent STM study however, no such effect occurred. Pairs of adjectives were no more likely to be forgotten when followed by a pair with a similar stimulus than when followed by an unrelated pair.

All the experiments described have been concerned with the question of whether classical LTM interference effects can be shown to occur in STM. The present study tries the opposite approach by taking a well-established phenomenon of STM and attempting to reproduce it in LTM.

The phenomenon selected was that demonstrated by Conrad (1964), who showed that items which are hard to discriminate when heard in noise are also hard to recall even though presented visually. A subsequent study by Baddeley (1966) of STM for word sequences showed that this was not a general similarity effect for although acoustic similarity led to a massive decrement in performance (72·5 per cent.), similarity of meaning had only a slight effect (6·3 per cent). Experiment I takes the material used in this study and attempts to produce a similar effect in LTM.

Experiment I

Method

Four lists of 10 words were used based on the sets employed by Baddeley (1966). List A comprised 10 acoustically similar words (man, cab, can, cad, cap, mad, max, mat, cat, map) and List B was a control list comprising acoustically relatively dissimilar words which were matched with List A for frequency of occurrence in the language using the Thorndike and Lorge (1944) norms (pit, few, cow, pen, sup, bar, day, hot, rig, bun). List C comprised 10 semantically similar adjectives (great, large, big, huge, broad, long, tall, fat, wide, high), and List D, adjectives of equal frequency of occurrence but with dissimilar meanings served as a control list (good, huge, hot, safe, thin, deep, strong, foul, old, late).

A separate group of subjects attempted to learn each of the four sequences using a learning procedure which was based as closely as possible on the technique employed in the previously described STM experiment (Baddeley, 1966). The words were presented by tape recorder at the rate of one per 3 sec. in the order given above,* and subjects were then allowed 40 sec. to write down in an answer booklet as much of the sequence as they could remember in the appropriate order. This procedure was repeated for a total of four trials, after which subjects spent 20 min. on an intervening task involving immedate memory for sequences of eight digits. They were then retested by being asked to write down as much of the 10-word list as they could in the appropriate order. Subjects were not warned in advance about the subsequent retest. As in the STM studies described previously, the relevant sets of words were written on cards which were visible throughout the experiment. This maximizes response availability and makes the test one of order rather than item learning. To prevent the use of position on the card as a cue, four cards were used, each with a different arrangement, and were frequently interchanged.

In order to ensure that subjects were able to hear the words accurately, the learning session was preceded by a listening test in which the 10 relevant words were read out in random order and the subject was allowed 5 sec. per word to copy them. Three subjects, all from Condition A were discarded for failing to score perfectly on the listening test.

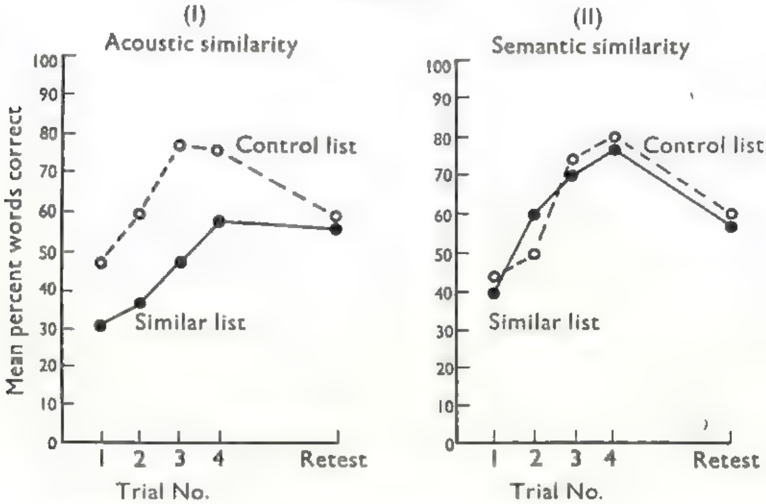
* While this is a slower rate of presentation than that used in the STM study (Baddeley, 1966), a subsequent experiment had shown that this is not a crucial difference. At a 3 sec. rate the effect of acoustic similarity on STM was greater than that of semantic similarity for 18 of the 20 subjects tested.

leaving 18, 17, 20 and 20 subjects in Conditions A, B, C and D respectively. The subjects were young servicemen and they were tested in groups of approximately 20.

Results

Performance was scored in terms of number of words reproduced in the appropriate position in the sequence and is shown in Figure 1. Since scores were not normally distributed they were analysed using non-parametric tests. Comparisons between conditions were made using the Mann-Whitney Test, and the significance of forgetting was estimated by comparing Trial 4 with Recall Test scores using the Wilcoxon Test. Unless otherwise stated two-tailed tests were used.

FIGURE 1



Learning and retention of 10-word sequences as a function of acoustic similarity (I) and semantic similarity (II).

Acoustic similarity. The similar list was harder than the control list on Trial 1 ($p < 0{\cdot}02$). This difference was maintained throughout learning but was only marginally significant on Trial 4 ($p < 0{\cdot}05$, one-tail). There was no difference between the two conditions in number of words retained. Significant forgetting occurred in the control list ($p < 0{\cdot}01$) but not in the acoustically similar list. The difference between the two in amount of forgetting was significant ($p < 0{\cdot}05$).

Semantic similarity. There were no significant differences between similar and control lists during learning or on the recall test. Both conditions showed a similar amount of forgetting and in both cases this was statistically significant ($p < 0{\cdot}02$).

Serial position effects. There were no marked differences between the four conditions, all of which showed the typical bow-shaped curve (McGeoch and Irion, 1952).

These results show therefore that the acoustically similar list differs from the other three lists in showing lower performance on all the learning trials. Unlike the other lists it shows no forgetting, so that performance after retest does not differ among the four groups. This result is puzzling for two reasons. First because in failing to show an effect of semantic similarity on LTM, it conflicts with previous experiments which have shown such an effect for both serial learning (Underwood and Goad, 1951) and paired associate learning (Underwood, 1951; Baddeley and Dale, 1966).

Secondly this result presents the paradox of a high level of performance being followed by a greater amount of forgetting than a low level, a result which is also in conflict with previous studies (Underwood, 1949*b*, p. 527).

One possible resolution of this paradox, however, is as follows:

In interpreting Experiment I we have assumed that the learning score reflects the gradual increase in material committed to LTM. While LTM is almost certainly *one* determinant of the learning score, however, it is not necessarily the only one. Since this score is based on a test which immediately follows presentation of the list, it seems reasonable to suppose that it is also affected by material which is in STM. On the other hand, the retest 20 min. later seems likely to depend only on LTM. The difference between these two scores may thus be determined by the amount of material in STM, and the forgetting shown in Experiment I may simply reflect the loss of this material. The one condition which does not show this loss is the acoustically similar list. Since STM for this material is known to be very poor (Baddeley, 1966), it seems reasonable to suppose that subjects in this condition would rely on STM as little as possible, and would thus show less forgetting between Trial 4 and the retest. This explanation of the results was tested in Experiment II.

Experiment II

Peterson and Peterson (1959) have shown that when a subject is prevented from rehearsing by an intervening task, there is a rapid decay of material in STM. Experiment II attempts to wipe out the influence of STM during the learning phase by interposing such a task between each presentation and test of the list. The hypothesis outlined above would predict that the semantically similar list would then behave like the acoustically similar list in showing a lower learning score, but no forgetting, while the acoustically similar list should be relatively unaffected by this attempt to wipe out the effects of STM.

Experiment II

Method

The hypothesis was tested using two lists from Experiment I, the semantically similar and the acoustically similar. Each list was learned by three groups of subjects using one of the following procedures. (X) Learning and testing were as in Experiment I. (Y) As in Experiment I except that a second task was interposed between each presentation of the list and each test. (Z) As in Y, except that the second task was interposed between each test and the following presentation. This condition was intended as a control for any general distribution of practice effects, without acting as a specific STM control.

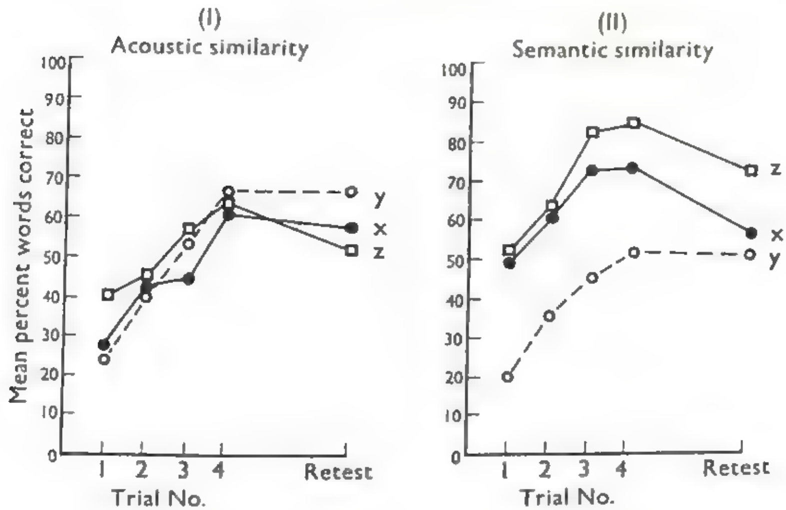
The intervening task was that used to fill the 20 min. separating the learning phase from the subsequent retest in both Experiment I and the present study. It involved the presentation of sequences of eight random digits at a rate of one per sec. After each sequence the subject was allowed 8 sec. to write down as many of the digits as he could recall in the order presented. In Condition Y and Z, each interposition comprised six eight-digit sequences. Apart from this, material and procedure were identical with Experiment I.

Housewives from the A.P.R.U. subject panel served as subjects and were tested in groups of approximately 20. Six subjects were discarded for failing the listening test, all from groups given the acoustically similar list. In the acoustically similar conditions this left groups of 20, 25 and 23 for groups X, Y and Z respectively, and groups of 17, 26 and 20 for the equivalent semantically similar conditions.

Results

The mean number of words reproduced in the correct position is shown in Figure 2. The results were analysed nonparametrically as in Experiment I.

FIGURE 2



Learning and retention of word sequences as a function of acoustic similarity (I), semantic similarity (II) and learning procedure. X = Standard procedure; Y = STM control between each presentation and test; Z = STM control between each test and the subsequent presentation.

Acoustically similar list. The STM control had no significant effect on performance during either the learning or recall phases, and neither X nor Y showed a significant amount of forgetting between Trial 4 and the Recall Test. Group Z however showed a significantly higher level of performance on Trial 1 ($p < 0{\cdot}02$), together with a significant amount of forgetting between Trial 4 and the Recall Test ($p < 0{\cdot}01$). Apart from the Trial 1 score, however, none of the differences between the groups was statistically significant.

Semantically similar list. The addition of an STM control reduced the learning score very considerably, the difference between X and Y was statistically highly significant on Trial 1 ($p < 0{\cdot}001$) and was still present on Trial 4 ($p < 0{\cdot}02$). As in Experiment I the straightforward semantically similar condition showed significant forgetting between Trial 4 and the Recall Test ($p < 0{\cdot}05$). This was significantly greater than the negligible loss shown by the STM control condition ($p < 0{\cdot}05$), the actual amount retained not being significantly different. Condition Z did not differ significantly from X at any point. There was significant forgetting in this condition ($p < 0{\cdot}05$) but the amount retained was significantly greater than in Condition Y ($p < 0{\cdot}02$).

It seems clear from these results that the conventional testing procedure confounds the effects of LTM and STM during the learning phase, and that this in turn may influence the subsequent recall score. This is particularly true using the present technique in which the subject can see his responses throughout the recall phase since this may help to transfer them from STM to LTM.

Performance in the Y Conditions suggests that it is possible to eliminate or at least minimize this STM effect, while the Z Conditions, where STM is if anything encouraged, show that the STM control has a specific effect which is not simply the result of distributed practice.

Experiment II suggests then that conventional testing procedures cannot be expected to give a satisfactory answer to our original question of the relative

importance of semantic and acoustic factors in LTM. Neither, however, does Experiment II itself since the appropriate control groups were not included. A third experiment was therefore performed using the STM control technique explored in Experiment II and the four lists from Experiment I. Finally, to eliminate any complication due to mishearing, visual presentation was used.

Experiment III

Method

A separate group of subjects learned each of the four lists used in Experiment I. As in Condition Y of Experiment II, however, each presentation of the list was followed immediately by a task involving STM for six sequences of eight digits. Words were presented by slide projector, they were visible for 3 sec. and the slide changeover time was approximately 2 sec. The intervening digit sequences were read out at a 1 sec. rate and subjects were allowed 8 sec. to write out each sequence. They were then allowed 1 min. to write out the 10-word sequence.

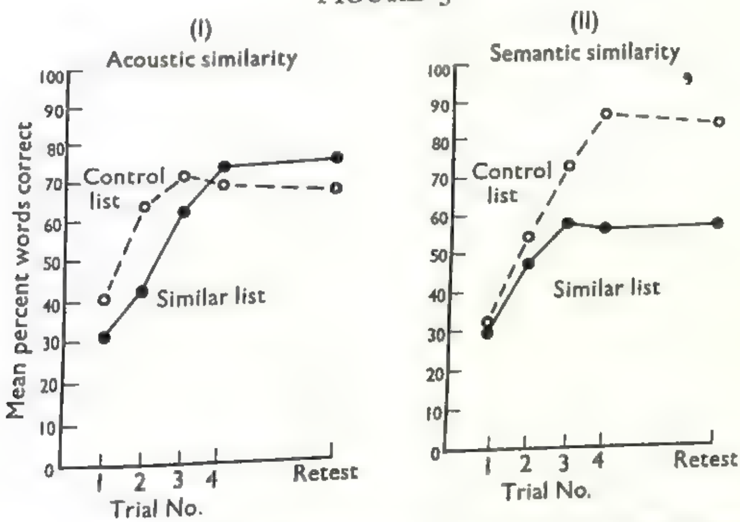
After four trials subjects performed a task involving 15 min. of self-paced digit copying. They then attempted to recall the word sequence. As in the previous experiments the relevant words were visible during learning, testing and the retest.

Mixed groups of men and women from the A.P.R.U. subject panel were tested. The acoustically similar list (A) was learned by 15 subjects, its control list (B) by 20, the semantically similar list (C) by 16 and its control (D) by 21 subjects.

Results

Performance on the four lists is shown in Figure 3.

FIGURE 3



Learning and retention of word sequences as a function of acoustic similarity (I) and semantic similarity (II) with STM controlled.

Acoustic similarity. There was a tendency for the similar list to be harder during early learning, though this was statistically significant only on Trial 2 ($p < 0.05$), and was reversed by Trial 4. Neither group showed any evidence of forgetting between Test 4 and the retest when performance on the acoustically similar list was not significantly different from the control list.

Semantic similarity. Although there was little difference on Trial 1, the semantically similar list showed slower learning, so that by Trial 4 scores were significantly higher on List D than List C ($p < 0.005$). Neither list showed any

forgetting and on the retest, performance on the semantically similar list was still reliably poorer than on the control list ($p < 0{\cdot}005$).

Discussion

The results of Experiment III indicate that the learning of word sequences was impaired by semantic similarity, and that in all groups what was learned was retained for at least 15 min. Together with the results of the previous experiments this suggests that LTM may be based either on the meanings of the words in the message or else on the sound of the words. This latter strategy, however, seems to involve an intermediate stage at which the material is in STM. This influence of STM is indicated both by higher scores during learning than are found on retest using the standard procedure, and by the difficulty of learning the list of semantically similar words when the influence of STM is minimized in the modified procedure used in Experiment III. It seems then that whereas STM relies very largely on acoustic coding and is relatively unaffected by the semantic content of the message to be stored (Conrad, 1964; Baddeley, 1966), LTM uses semantic coding extensively (Underwood, 1951; Underwood and Goad, 1951; Baddeley and Dale, 1966), though not exclusively. The message *may* be learned on the basis of acoustic coding as in Experiment I, though in this case STM plays an important role, as is shown in Experiments II and III. This conclusion is supported by the results of Anderson (1965), who has performed a parallel series of experiments on the learning and retention of letter sequences. As in Experiment I she finds that an acoustically similar sequence shows poorer learning but better retention than a control sequence, and as in Experiment III this difference disappears when an STM control is used.

It seems clear then that the extreme vulnerability of STM to the effects of acoustic similarity is not shared by LTM. On the other hand Experiment III agrees with Underwood (1951) and Underwood and Goad (1951) in suggesting that LTM *is* affected by semantic similarity in the learning stage, and with Underwood (1951) in finding no effect of semantic similarity on forgetting.

The effect of the STM control in both this study and Anderson's emphasises the role that STM may play in LTM experiments. The probable interaction of STM and LTM has been discussed by Murdock (1963), though the intentional use of an STM control to isolate LTM effects does not appear to be common practice.

The experiments described have shown that material which is difficult for STM is not necessarily difficult when used in an LTM study, for while in STM acoustic cues appear to be of crucial importance, in LTM semantic cues may be used instead. What constitutes the crucial difference between LTM and STM, whether time in store, number of repetitions or the coding system employed, however, remains to be seen.

References

Anderson, N. S. (1965). Effects of acoustic similarity in serial learning and retention. *Paper read Psychonomics Society, Chicago.*

Baddeley, A. D. (1966). Short-term memory for word sequences as a function of acoustic, semantic and formal similarity. *Quart. J. exp. Psychol.*, **18**, 362–5.

Baddeley, A. D., and Dale, H. C. A. (1966). The effect of semantic similarity on retroactive interference in long- and short-term memory. *J. verb. Learn. verb. Behav.* (in press).

Conrad, R. (1964). Acoustic confusions in immediate memory. *Brit. J. Psychol.*, **55**, 75–84.

Conrad, R. (1966). Interference or decay over short retention intervals? *J. verb. Learn. verb. Behav.* (in press).

KEPPEL, G., and UNDERWOOD, B. J. (1962). Proactive inhibition in short-term retention of single items. *J. verb. Learn. verb. Behav.*, **1**, 153–61.

LOESS, H. (1964). Proactive inhibition in short-term memory. *J. verb. Learn. verb. Behav.*, **3**, 362–8.

McGEOCH, J. A., and IRION, A. L. (1952). *The Psychology of Human Learning*. New York: Longmans Green.

MELTON, A. W. (1963). Implication of short-term memory for a general theory of memory. *J. verb. Learn. verb. Behav.*, **2**, 1–21.

MURDOCK, B. B. (1963). Short-term memory and paired-associate learning. *J. verb. Learn. verb. Behav.*, **2**, 320–8.

PETERSON, L. R., and GENTILE, A. (1965). Proactive interference as a function of time between tests. *J. exp. Psychol.*, **70**, 473–8.

PETERSON, L. R., and PETERSON, M. J. (1959). Short-term retention of individual verbal items. *J. exp. Psychol.*, **58**, 193–8.

POSTMAN, L. (1964). Immediate memory and incidental learning. In MELTON, A. W. (Ed.), *Categories of Human Learning*. New York: Academic Press, pp. 145–201.

THORNDIKE, E. L., and LORGE, I. (1944). *The Teacher's Word Book of* 30,000 *Words*. New York: Teachers College, Columbia University.

UNDERWOOD, B. J. (1949*a*). Proactive inhibition as a function of time and degree of prior learning. *J. exp. Psychol.*, **39**, 24–34.

UNDERWOOD, B. J. (1949*b*). *Experimental Psychology*. New York: Appleton Century Crofts.

UNDERWOOD, B. J. (1951). Studies of distributed practice: II. Learning and retention of paired adjective lists with two levels of intra-list similarity. *J. exp. Psychol.*, **42**, 153–61.

UNDERWOOD, B. J., and GOAD, D. (1951). Studies of distributed practice: I. The influence of intra-list similarity in serial learning. *J. exp. Psychol.*, **42**, 125–34.

Manuscript received 17th March, 1966.

THE EFFECT OF INFORMATION CONTENT UPON THE PERCEPTION AND AFTER-EFFECTS OF A ROTATING FIELD

BY

N. F. DIXON and LINDA MEISELS

From University College, London, and Grinnell College, Grinnell, Iowa, U.S.A.

From an investigation of movement after-effects induced by a rotating field, it seems that the information content of the inspection field is an important determinant of the subsequent movement after-effects (M.A.E.). This finding, considered in conjunction with phenomena evoked during perception of high information content and highly redundant fields, is discussed in connection with Anstis and Gregory's (1965) work on the role of retinal stimulation in the production of M.A.E.s.

Introduction

If the visual after-effects of perceived motion depend upon the retina/image rather than head/eye signalling system (Anstis and Gregory, 1965) it would seem reasonable to assume that the information content of the field would be an important determinant of the extent and duration of the after-effect. A second argument for supposing this to be the case stems from recent work by Barlow, Hill and Levick (1964) on the rabbit retina. These writers found that regular repetitive stimulation brought about a reduction in firing of certain movement sensitive receptor units. If similar units exist in the human retina, it is again reasonable to assume that movement after-effects (M.A.E.s) would be greater when prior stimulation by a moving field is randomly rather than regularly distributed in time.

It was with these considerations in mind that we have been examining the hypothesis that a high information content field would evoke movement after-effects in excess of those generated by an otherwise equivalent field having high internal redundancy.

Apparatus and method

The stimulus fields consisted of discs, 15·2 cm. in diameter, rotated slowly in the subject's fronto-parallel plane. For the high information content field (disc "A"), the display consisted of an equal number of quasi randomly distributed black and white squares. Each of the 21,000 squares in this field, would, by itself, subtend an angle of 3′ at a distance of 1 metre from the subject. The other, highly redundant, display (disc "B") consisted of a chequer board pattern of black and white squares in which each square subtended an angle of 6′ at 1 metre. The reason for using larger squares in disc B was to equate as nearly as possible the number of black white changes produced by the two discs in unit time at any given speed of rotation on any given area of the retina.

At 1 metre distance both discs subtended an angle of 8° 42′.

For the first experiment 11 subjects were required to fixate either disc "A" or disc "B" for a given period of time and then transfer their gaze to a plain grey disc of the same size and at the same distance from them. This grey disc, the after field, was mounted on the output end of a Selsyn link, so that the subject, by turning a knob, could rotate it in either direction as far and as fast as he might wish. In fact, the subjects were instructed to manipulate the grey field in such a way as to make it appear stationary after viewing the test field. The extent to which the subjects rotated the after field was registered on a dial visible only to the experimenter.

Experiment I

Procedure

For the first experiment half the subjects were presented first with disc "A" and secondly with disc "B." The remaining subjects were given the discs in the reverse order. For all subjects, rotation speed was fixed at 2 r.p.m. and the distance of both the inspection and test fields at 2 metres.

For this same experiment each subject was tested once for each of three inspection times for each disc. These were 5, 10 and 20 sec., the order of treatments being varied across subjects. After inspecting one of the two test fields for the appropriate period, subjects transferred their gaze to the test field and "demonstrated" the after-effect as mentioned above. The remaining experiments described in this paper were devoted to a more detailed analysis of the M.A.E.s (movement after-effects) induced by disc "A," by varying such parameters as rotation speed and viewing distance of the inspection figure.

Results

Table I and Figure 1 show the relationships between viewing time and after-effect for the two sorts of disc. As might be expected, the duration and extent of the movement after-effect is largely determined by the viewing time. These data, however, also support our hypothesis that the after-effect is in part a function of the information content of the inspection field. In respect of both duration and angular displacement, the effect evoked by disc "A" exceeded that of disc "B." For both measures the difference between the discs was significant at the 0·5 per cent. level of confidence ($t = 3{\cdot}45$ and $3{\cdot}23$ respectively).

TABLE I

Experiment I

Viewing distance: 2 m.
Viewing time: *Variable*
Velocity: 2 r.p.m.

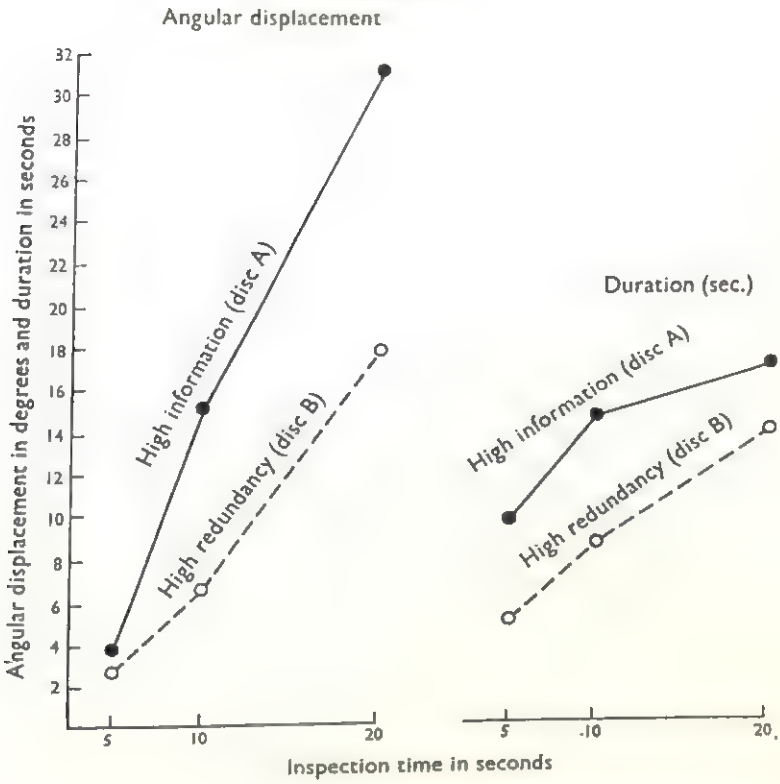
	Random disc						Patterned disc					
	5 sec.		10 sec.		20 sec.		5 sec.		10 sec.		20 sec.	
Subject	Ang°.	Dur. sec.	Ang°.	Dur. sec.	Ang°.	Dur. sec.	Ang°.	Dur. sec.	Ang°.	Dur. sec.	Ang°.	Dur. sec.
1	1	5	11	17	30	18·5	9	5	1	8	27	10
2	0	4	6	7	14	13	0	8	1	6	4	6
3	2	4	8	8	43	22	0	3·5	4	7	16	29
4	4	9	12	18	52	26	0	0	5	16	48	19
5	4	10	14	14	27	11	1	5	14	10	14	10
6	0	5	14	6	8	5	9	6	7	6	5	5
7	0	9	18	17	11	18	1	5·5	6	8	1	9
8	22	35	58	37	78	30	10	15	27	22	35	28
9	0	2·5	11	6	44	11	1	4	9	8	24	7
10	7	14	11	21	10	10	0	4	0	7·5	9	10
11	1	9	2	11	28	27	0	0	0	0	19	23
$\overline{X}$	3·73	9·68	15·0	14·73	31·36	17·41	2·82	5·09	6·73	8·95	18·36	14·18

(To control against the possibility that the larger size of the squares in disc "B" might be contributing to the relative reduction of M.A.E. for the redundant field, an additional three subjects were run in which the size of squares in the high information content field *exceeded* those in the high redundancy field. The results suggest that far from decreasing the M.A.E. differential as between the random and patterned fields, this change in stimulus size evoked an even larger M.A.E. for the high information content field.

A further study is being conducted to examine the exact relationship between stimulus size and M.A.E. for the two sorts of field.)

FIGURE 1

Experiment 1



A comparison between the M.A.E.s of high and low information content fields as a function of inspection time (11 subjects).

Experiment 2

The very large and long lasting M.A.E.s evoked by the high information content field raised the question as to how M.A.E.s would respond to changes in information rate induced by altering the rotation speed.

Exactly the same procedure was followed as in Experiment 1, except that now with viewing time held constant, subjects were required to inspect disc "A" under four different rotation speeds—1, 2, 4 and 6 r.p.m.

The results are shown in Table II and Figure 2.

These data suggest a curvilinear relationship between stimulus velocity, and the M.A.E. related measures, of duration and angular displacement

Nine of the 11 subjects showed larger M.A.E.s when the disc was rotating at 2 to 4 r.p.m. than when it was travelling at higher or lower speeds. It is perhaps noteworthy that this optimum speed of around 18° per sec. is in close agreement with the optimum velocity for "movement masking" found by Grindley and Townsend (1965). According to evidence referred to by these authors, this velocity also approximates to that which produces maximal disturbance in the superior colliculus of the rabbit.

A *post hoc* finding, illustrative of the strength of the M.A.E. induced by disc "A," was that for nine of the 11 subjects the duration of the M.A.E. significantly exceeded that of the inspection period ($t = 3{\cdot}00$; $p < 0{\cdot}02$ two-tail).

TABLE II

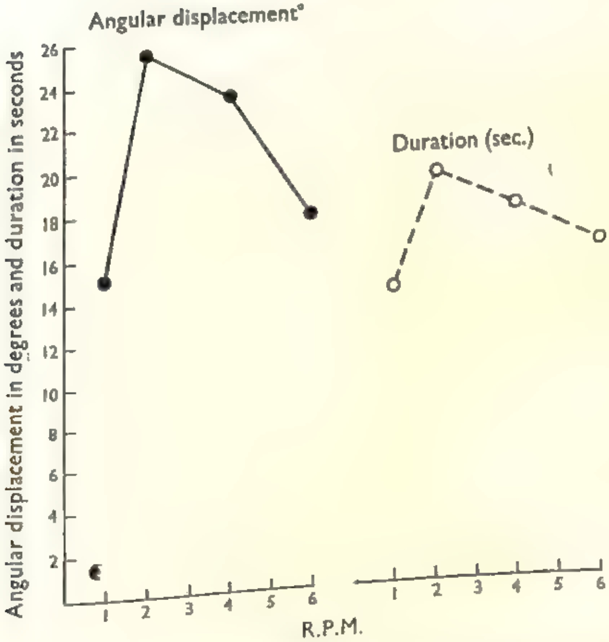
EXPERIMENT 2 (DISC "A")

Viewing distance: 2 m.
Viewing time: 10 sec.
Velocity: *Variable*

Subject	1 *r.p.m.*		2 *r.p.m.*		4 *r.p.m.*		6 *r.p.m.*	
	Ang°.	*Dur. sec.*	*Ang°.*	*Dur. sec.*	*Ang°.*	*Dur. sec.*	*Ang°.*	*Dur. sec.*
1	3	7	14	6	5	4	12	3
2	0	5	0	0	4	6	15	9
3	0	0	12	28	22	33	11	24
4	50	21	99	25	110	21·5	59	25
5	9	18·5	29	22·5	15	19	24	18
6	41	23·5	51	28·5	72	32	51	27·5
7	5	16	10	22·5	9	23	9	16
8	26	24	33	26	6	29	2	11
9	6	9	1	11·5	5	13	10	12
10	22	16·5	14	20	11	17	2	12
11	4	15	19	23	0	0	1	18
$\overline{X}$	15·1	14·1	25·6	19·4	23·5	17·9	17·8	15·9

FIGURE 2

Experiment 2



M.A.E.s induced by a high information content field (disc "A") as a function of rotation speed (11 subjects).

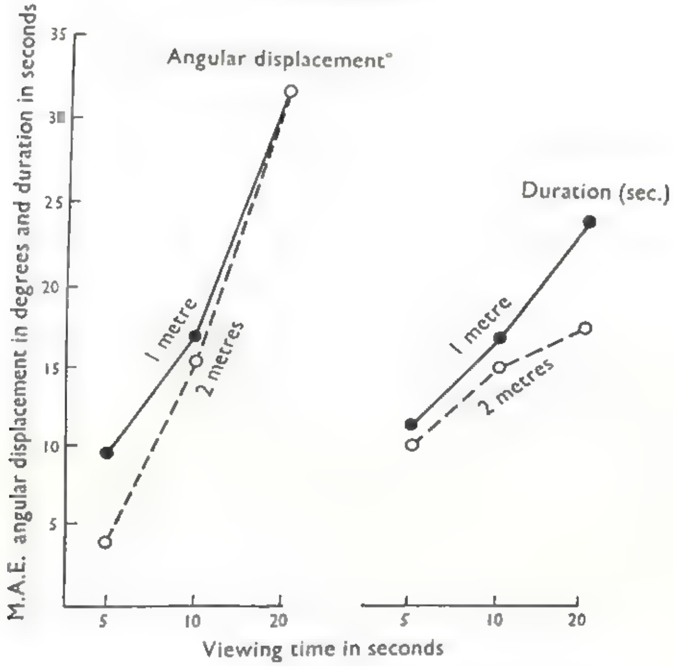
Experiment 3

Distance of inspection and test fields

A problem raised by the results from the preceding experiment is the relationship between the M.A.E. and channel capacity of the retina. For this reason trials were run in which the M.A.E.s induced by a field at 2 metres, could be compared with those induced by the same field at 1 metre. In the latter case there are, of course, considerably more retinal units carrying the same amount of information as hitherto.

FIGURE 3

Experiment 3



M.A.E.s for a rotating high information content field as a function of inspection time and viewing distance (11 subjects).

As these curves show, there is a small but insignificant increase in the M.A.E. when the inspection and test fields are at half the distance. The number of retinal receptors affected by each black/white change does not therefore appear to be a critical variable in production of the M.A.E. unless it be assumed that, increasing the total retinal area stimulated, exactly compensates for the reduction in information transmitted by each individual receptor.

Discussion

Before discussing the implications of these results it is, perhaps, worth considering certain other phenomena associated with perception of these rotating fields.

The high information content field (disc "A")

While in some respects a version of the old waterfall illusion studied in considerable detail by Wohlgemuth (1911), the M.A.E. induced by this field far exceeds in severity, longevity and velocity any other visual after-effect known to the authors

of this paper. It appears to have two components—a viscid distortion or "creeping" of any surface looked at, coupled with a clear-cut rotation of any form or object in the test field. After prolonged inspection of the rotating stimulus field the after-effect is sudden and violent, the stationary test appearing to "uncoil" at a very much higher velocity than that at which the inspection field had been moving.

During inspection of the rotating stimulus field, changes in the part of the display fixated appear to have little influence upon the subsequent after-effect, even though the same group of retinal receptors may be swept from left to right at one moment, and in the reverse direction immediately afterwards. In other words, the after-effect appears to be as much dependent upon the average firing order over a period of time of a large number of retinal receptors as upon the total number of receptors firing in the appropriate order at any one time. Whether fixation is stationary or moving the after-effect is always torsional and contrary in direction to that of the total stimulus field.

Those regions of a subsequent test surface, upon which the after-images of small stationary areas in the inspection field are superimposed,* remain stationary despite unimpaired apparent movement of the surrounding test field. While this may seem to imply a retinal locus for the effect, there is not, in fact, any induced contrary motion in the after image such as might be expected were the M.A.E. entirely peripheral in origin.

Following a fixation point attached to the inducing field does not cause the M.A.E. to disappear. This in itself is not necessarily incompatible with Anstis and Gregory's finding of no after-effect when the eye is following a moving display of vertical stripes, for in the case of a rotating field there remains a considerable degree of retinal image displacement.

As for the spiral after-effect (Wohlgemuth, 1911; Spigel, 1962), the potential for an after-effect induced by a random distribution of black and white squares, is preserved during a dark period interposed between the inspection and test fields. The relationship, *Experiment* 4, between dark period and extent of the M.A.E. is shown by the following curve. (See Fig. 4).

It appears that after 10 sec. in the dark the subsequent M.A.E. from inspection of disc "A" for 10 sec. is hardly less than that occurring when the M.A.E. is tested immediately following fixation of the inducing field. In other words, the return to a *status quo* of the visual system appears to require some visual input. The impression received is that the visible static test surface "releases" the after-effect. This preservation of the M.A.E. potential during darkness suggests an explanation in terms of storage in some reverberatory system rather than one in terms of either field effects or neural adaptation.

One final point about the M.A.E. induced by this high information content rotating field is that though very greatly reduced, it is still experienced by most subjects when the inspection field is presented to one eye and the test field to the other.

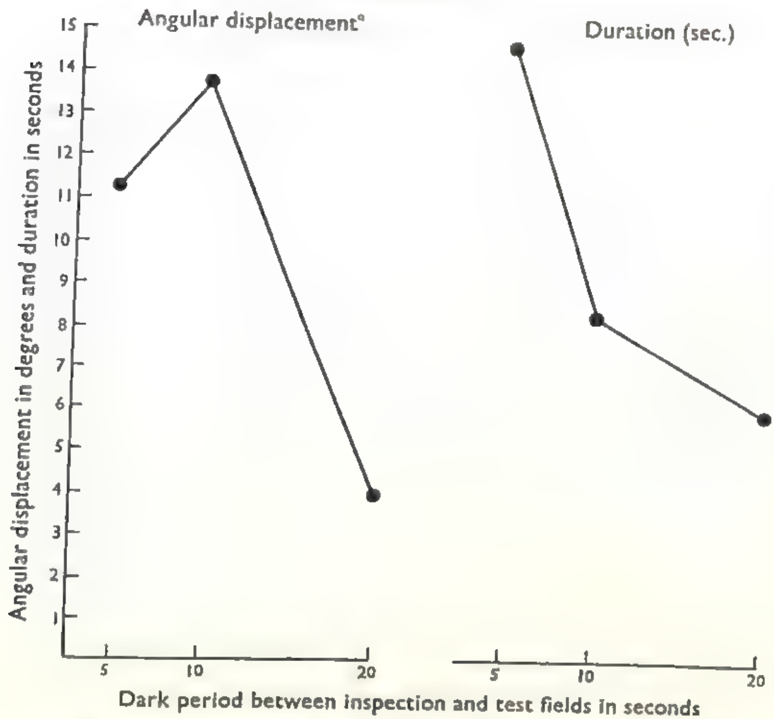
Rotating fields with high internal redundancy (e.g. disc "B")

In addition to the field (disc "B") used in the above mentioned experiment, we have experimented informally with several other patterned surfaces, e.g. small black dots on a white ground, and a fine grid of black lines on a white ground. All these fields have two features in common with the disc "B" field (black and white squares). The first is that the M.A.E. they evoke is less than that induced by the high information

* Small opaque coloured discs were held magnetically upon a sheet of glass, through which the inspection field could be viewed. The negative after-images of these discs were plainly visible on the subsequent test field.

FIGURE 4

Experiment 4



The effect of a dark period interposed between inspection of a rotating high information content field and subsequent M.A.E.s (6 subjects).

content field (disc "A"). The second is that at some rotational speed between 2 and 20 r.p.m. their surface structure appears to change in a characteristic way. The first change is that shimmering white bands begin to appear moving back and forth and up and down across the total field. This effect, which resembles the "moiré" and "streaming" phenomena described by MacKay (1961), is followed by the appearance of a large grey cross centred about the middle of the disc and rotating with it. An analysis of the latter phenomenon suggests that at a velocity well below the c.f.f. the stimulus field becomes structured along the lines of adjacent similar elements having parallel loci about the centre of the field. Thus in the black and white chequer board field the cross lies through the diagonals of those black squares which together constitute a diameter of the disc. Similarly, for the field containing black dots on a white ground the cross appears along the two widest, continuously white, diameters.

In every case the emergence of the cross is accompanied by a blurring and recession of the remainder of the rotating field. These phenomena, which have a physical basis in the sense that the retina integrates in time *longer* lasting impressions of continuous whiteness or blackness while suppressing those of shorter duration, serve to resolve a rotating field into a simpler figure/ground relationship.

In this latter respect they have their counterpart in the high information field. This too, though at much slower speeds, and for less obvious reasons, tends to resolve itself into a number of simpler forms upon an amorphous ground. For this type of field the patterning imposed by increased information rate consists at first of vague

swirling shapes which at higher speeds show transformation into contoured ovals or leaf-like figures.

While it is beyond the scope of this paper to assess the precise significance of these phenomena for the M.A.E. described earlier, it can be said that—*if* the complexity of patterning evoked by the rotating field is indicative of the simplest possible recoding, or redundancy stripping of the visual information transmitted, these phenomena tend to confirm the relationship between information rate and M.A.E. shown above.

In other words, the higher the information rate the more complex the recoding that is necessary for the preservation of structure and the more complex the recoding the larger the M.A.E.

Once integration has occurred the number of movement receptors required at any one moment to transmit rotatory movement of a cross must be very much less in number than those mediating perception of a complex and unblurred field of random elements. *If* M.A.E. displacement and duration are a function of the number of receptors differentially involved our results become explicable.

Summary and Conclusions

In the course of investigating a hypothesis suggested by the work of Anstis and Gregory (1965), data were obtained which suggest:—

(1) M.A.E.s evoked by a rotating field are positively related to:—

(*a*) Information content of the field

(*b*) Inspection time

(*c*) Velocity up to a certain maximum rate beyond which they decline.

(2) The potential for an M.A.E. can be preserved during an intervening dark period.

(3) The duration of the M.A.E. tends to exceed that of the inspection time.

(4) When the angle subtended by the total field and that subtended by the individual elements of which it is composed are altered together, there is very little effect on the M.A.E.

(5) The M.A.E. does not occur for what were stationary areas within the rotating field.

(6) Following a fixation point which moves with the field does not abolish the M.A.E.

(7) Though diminished in extent the M.A.E. shows interocular transfer.

(8) A high information content field (randomly distributed black and white squares) resolves itself into oval or leaf-like forms. Regularly patterned fields resolve themselves into cross-like forms along diameters of identical elements. The complexity of these transformations appear related to extent of M.A.E. evoked.

In the main these results are consistent with the view that M.A.E.s depend upon movement signalled by the retina/image rather than head/eye signalling system.

There are, however, at least three senses in which the movement of an image across the retina may be described (a) in terms of a relative displacement of the whole image in which events at the boundary of the enclosing contour will be of prime importance; (b) in terms of the number of on/off events induced within the area covered by the image as it moves across; and (c) in terms of the novelty or unpredictability of each of these on/off events which together account for the total impression of movement.

From our results it would seem that it is the last of these aspects of signalled movement which determines the extent and duration of an M.A.E.

One entirely speculative explanation of this effect might be that, as has been shown for certain sub-human species, the human visual system includes movement sensitive

units which habituate to highly redundant repetitive stimulation. This explanation further assumes that withdrawal of input from a habituated system results in less after-effect, because at the moment of withdrawal the input from habituated and non-stimulated units would be more nearly equal than that from non-habituated and non-stimulated units. The possible validity of such an explanation and site of such habituation or adaptation as may be responsible for these phenomena are questions which must perforce wait upon further research, as must also the relationship, if any, between phenomena evoked during perception of a rotating field and those which occur as an after-effect of the same stimulation.

We are indebted to Mr. Andrew Szmidla for the high information content fields used in these experiments.

References

Anstis, S. M., and Gregory, R. L. (1965). The after-effect of seen motion: the role of retinal stimulation and of eye movements. *Quart. J. exp. Psychol.*, **17**, 173–4.

Barlow, H. B., Hill, R. M., and Levick, W. R. (1964). Retinal ganglion cells responding selectively to direction and speed of image motion in the rabbit. *J. Physiol.*, **173**, 377–407.

Grindley, G. C., and Townsend, V. (1965). Binocular masking induced by a moving object. *Quart. J. exp. Psychol.*, **17**, 97–109.

Mackay, D. M. (1961). Interactive processes in visual perception. In Rosenblith, W. A. (Ed.), *Sensory Communication*, Cambridge, Mass.: M.I.T. Press.

Spigel, I. M. (1962). Contour absence as a critical factor in the inhibition of the decay of a movement after-effect. *J. Psychol.*, **54**, 221–8.

Wohlgemuth, A. (1911). On the after effect of seen movement. *Brit. J. Psychol., Monogr.*, **1**, 1–117.

Manuscript received 12th March, 1966.

FURTHER EXPERIMENTS ON MOVEMENT MASKING

BY

G. C. GRINDLEY and VALERIE TOWNSEND

From the Psychological Laboratory, Cambridge

Voluntary attention to one of two static objects in the peripheral field of one eye makes this object *more* liable to masking by a moving object in the corresponding area of the field of the other eye (Experiment 1).

Positive after images (and probably negative after images) are subject to (binocular) movement masking (Experiment 2).

Movement masking can occur in the field of either eye, but with the displays so far tried the inhibitory influence of a moving object is less in the field of the eye to which it is shown than in the field of the other eye (Experiment 3).

Introduction

In earlier experiments (Grindley and Townsend, 1965) it was found that a moving object in a part of the visual field of one eye could obliterate temporarily a static object in the same or neighbouring parts of the field of the other eye. There was evidence that the frequency with which such masking occurs is affected by factors such as the velocity of the moving object and the part of the visual field concerned. The frequency of the masking was much greater in the periphery than in the centre of the field. The present paper is an extension of the earlier one and it is intended to investigate:—

(1) The effect of directing voluntary attention to the static object.

(2) The question of whether binocular masking of the kind described also applies to after images.

(3) Whether similar masking is observed when the static and moving objects are in the field of the same eye.

Experiment 1

The effect of voluntary attention

Voluntary attention to an object is usually regarded as something which makes perception of that object clearer; and in the case of an object in the periphery of the visual field there is some experimental evidence for this view (Helmholtz, 1925; Grindley, 1931; Averbach and Sperling, 1960). It was therefore thought possible that attention directed to a static object in the field of one eye might make the object less liable to masking by a moving object in the field of the other.

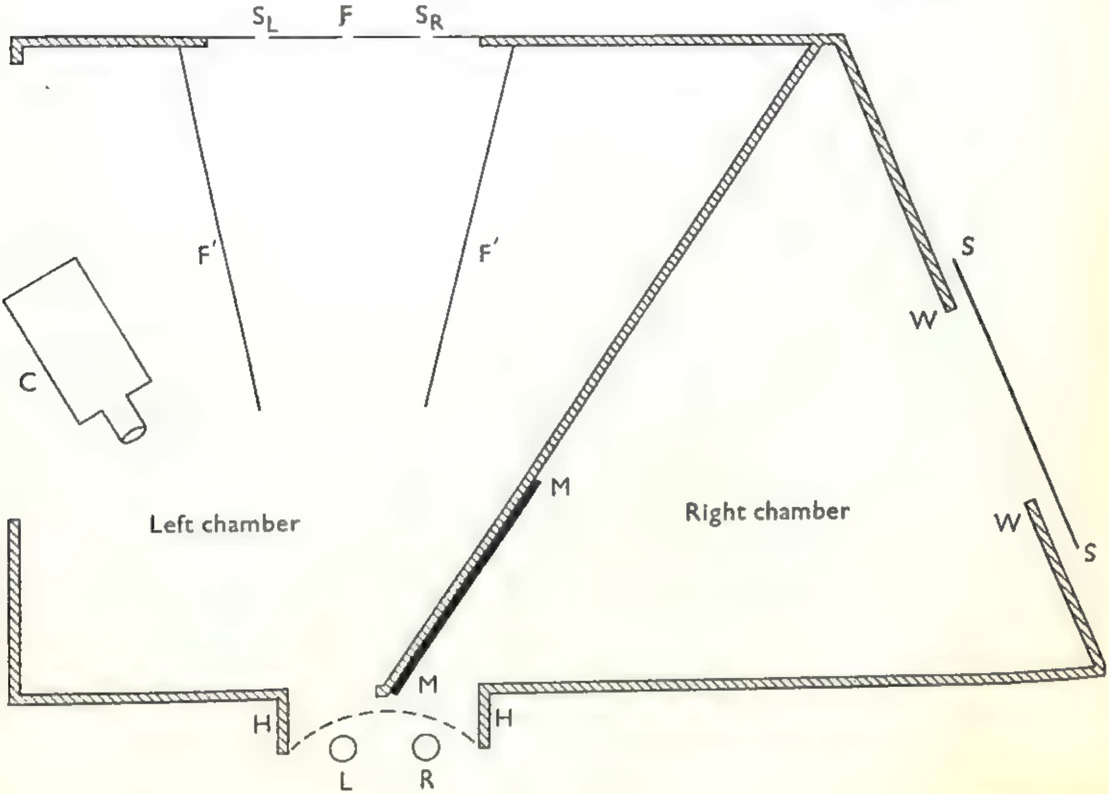
One of the main difficulties in the design of an experiment to test this is that movement masking is extremely sensitive to small eye movements (cf. Grindley and Townsend, 1965, p. 101). We therefore arranged the apparatus so that to the eye with the static field of view there was a fixation point with *two* objects, one on each side. Any small involuntary movements of this eye when the subject was asked to attend to one object would therefore produce equal retinal movements of both objects.

Apparatus

The apparatus is shown in plan in Figure 1: it is a modification of that described in our earlier paper. The subject sat with his head against a head rest (HH) and his left eye (L) viewed a fixation spot (F) on either side of which were two spots of light (S_L and

S_R). An opaque funnel F′F′ reduced the amount of stray light falling on the front of the display. The subject's left eye (L), which was illuminated by fluorescent lamps not shown in the diagram, could be watched through a telescope (C). The display to each eye is shown in Figure 2.

FIGURE 1



Plan of apparatus.

With his right eye (R) the subject viewed, through a mirror (MM), the image of a screen (SS) which could be caused to slide upwards or downwards at a controlled speed. There was an illuminated slit (S′S′ in Fig. 2) across this screen, and the right chamber was otherwise dark, so that what the subject saw with his right eye was a horizontal slit moving upwards or downwards on an otherwise dark field.

The optical pathway from either eye to the corresponding display was about 20 in. The spots S_L and S_R each subtended 0·75° visual angle and they were at 9° from the small fixation mark F. The width of the slit S′S′ also subtended an angle of 0·75° visual angle, and this slit moved upwards or downwards at 11° per sec. The window WW subtended about 24° × 22°, so that the slit was visible for about 2·0 sec. (1 sec. before crossing the spots and 1 sec. afterwards). The luminance of the spots S_L and S_R, of the fixation mark F, and of the moving slit S′S′ were all set at 10 ft. L.

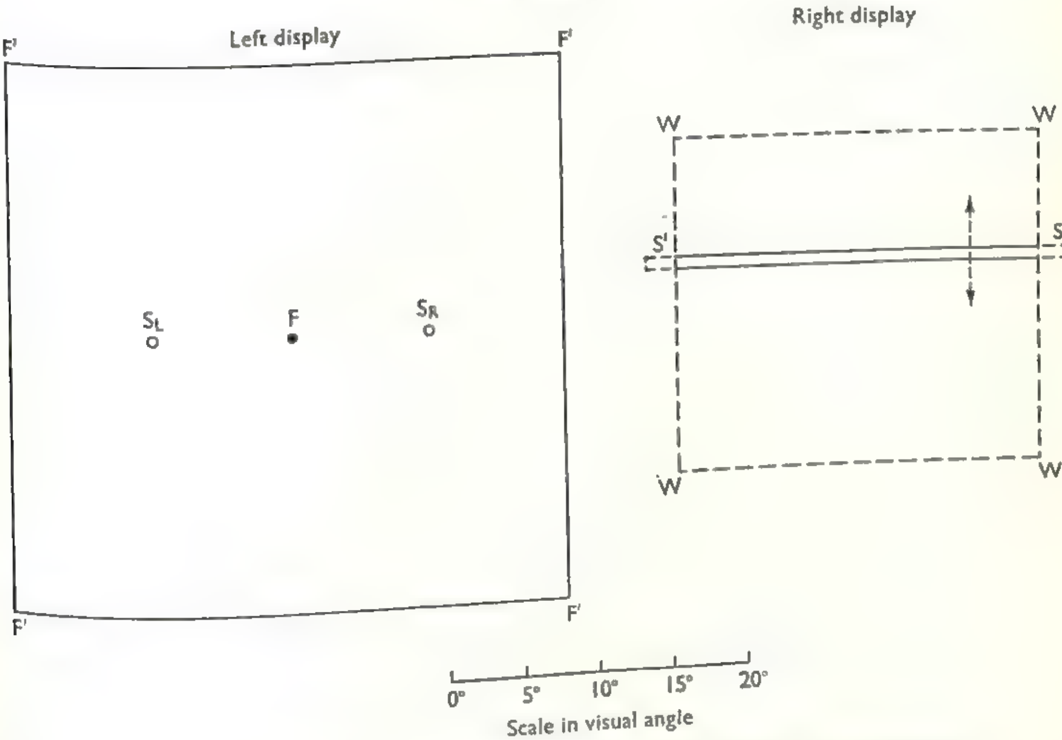
Procedure

Sixteen members of the Department were used for the experiment. At each trial the experimenter gave a warning signal of "ready" and the subject placed his head against the rest. The experimenter told the subject to "attend right" (or "left") or did not mention attention. The subject was then told to "fix" and he fixated on the small fixation point (F). At the same moment the experimenter started the motor and the slit appeared at the top (or bottom) of the window about 2 sec. later, crossing the spots about 1 sec. after this and passing from view 1 sec. later. After some practice trials each subject had 10 trials with no instructions about attention. This was followed by 20 trials in 10 of which the

instructions were to "attend right" and in the other 10 to "attend left." These trials were given in random order. Each trial lasted 7 sec. (i.e. until the slit had disappeared from view for about 3 sec.) and the subject only fixated for that time.

After each trial the subjects were asked which spot, if either, disappeared, went first, was back first, and which went for the longest time. The first five subjects reported their results verbally, the remaining 11 drew them diagrammatically after each trial. The subjects and the experimenters found the drawing method much the more satisfactory. Any trial in which there were detectable eye movements was ignored and repeated later.

FIGURE 2



Only the part of the slit shown in solid lines was visible to the subject.

Results

The conditions of the experiment were very favourable for movement masking. In 89 per cent. of all the trials one or both of the spots was reported to have disappeared during the 2 sec. when the slit was crossing the window. (In the remaining 11 per cent. of trials the subject reported either that there was no disappearance during this period, or that he was not certain what had happened; and these negative results were ignored).

The number of cases in which only one spot disappeared (under each of the three instructions) is shown in Table I. Similarly, when both spots disappeared, the cases in which either went first, or in which the disappearances were reported as simultaneous are shown in Table II. Table III is a combination of Tables I and II.

Inspection of these tables suggests:—

(*a*) that there was a strong tendency for the left spot to disappear before the right under all conditions of instruction. We did not explore the reasons for this bias.

TABLE I / TABLE II

Instructions	*(Only one disappeared)* *Left*	*Right*	*(Both disappeared)* *Left first*	*Simultaneous*	*Right first*
No Attention ..	18	2	47	72	12
Attention Left	25	3	59	42	5
Attention Right	13	20	33	60	23

TABLE III
(Tables I and II combined)

Instructions	*Left first or only*	*Simultaneous*	*Right first or only*
No Attention ..	65	72	14
Attention Left ..	84	42	8
Attention Right	46	60	43

(*b*) that voluntary attention to one spot makes it *more* liable to "movement masking," i.e. more liable to be the only spot which disappears, or to be the one which disappears first. This, of course, is quite contrary to what we expected, but is confirmed by looking at the results in a different way, i.e. by considering the results of each individual subject. Of the 16 subjects, 12 showed a majority of trials in which the spot attended to was the first (or only one) to disappear, three showed the opposite effect, and one showed no effect. The evidence seems convincing; and a brief comment will be made in the Discussion.

Experiment 2

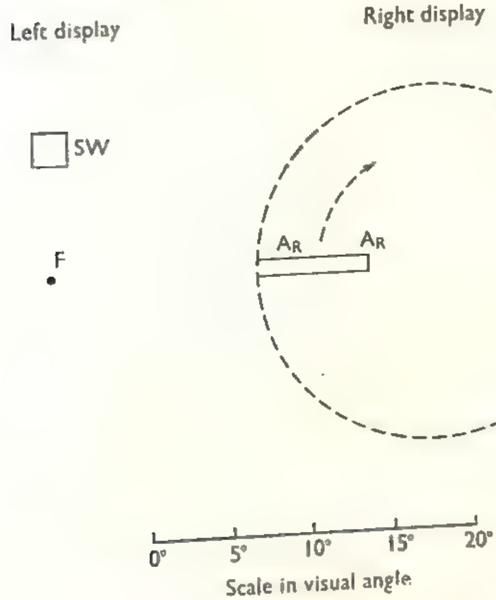
The effect of movement masking on after images

The previous experiments and those described in our earlier paper dealt with the masking of objects which were physically present. It was thought that it might be of interest to find whether similar effects occurred with after images. Certain difficulties were encountered in trying to find a satisfactory technique. If the eye is very strongly stimulated a fairly stable after image (positive or negative) can be produced and observed for as long as 20 min. (Brindley, 1962). But towards the end of its life such an after image becomes unstable, disappearing and reappearing at irregular intervals. We tried various durations and intensities of stimulation (including xenon lamp discharges) in an attempt to obtain a fairly clear and stable after image but not one that would persist so long as to make the experiment too laborious.

Apparatus

The apparatus was a modification of that shown in Figure 1. The telescope (C) was removed and the left chamber was completely darkened. The fixation spot F was as before, but instead of the spots S_L and S_R there was a square window SW subtending an angle of 2° with its centre 8° visual angle above F. The displays to the two eyes are shown in Figure 3. The luminance of the square was about 100 ft. L. and it was switched on for 7 sec. at each trial. It will be seen that with this apparatus it was possible to study only positive after images.

FIGURE 3



In the right display only the arm (shown in solid lines) was visible to the subject.

In the field of the right eye a black disc on which there was a white arm A_RA_R was substituted for the screen SS. The radius of the disc subtended an angle of 11° and the width of the arm subtended 1°. This disc could be made to rotate at 12 r.p.m. After a rough measurement of each subject's heterophoria, the disc was adjusted to such a position that its centre as viewed by the right eye usually coincided with the fixation spot F as seen by the left eye. The disc and its black surround were both illuminated from the front giving a luminance of 2 ft. L. for the white arm and about 0·25 ft. L. for the black disc and its surround. Thus when the disc was caused to rotate the white arm, as seen by the right eye appeared to cross the after image of the bright square which had been seen by the left eye. The illumination to the disc was left on even when the arm was not rotating. A recorder enabled the subject to register the moments when after images appeared or disappeared.

Procedure

The subject was told to fixate the fixation spot F and the square was then illuminated for 7 sec. The subject continued fixating and pressed a switch when a fairly clear after image appeared, releasing the switch when the after image disappeared. Nine subjects each had 10 trials, and in half of these the experimenter started the disc rotating as soon as the subject registered the appearance of a clear after image. In these trials the white arm appeared to cross the after image after 1·25 sec. In the other half of the trials (which were given alternately) the disc was stationary. Thus these trials acted as a check on the stability of the after image when it was not interfered with by movement masking.

Results

The results for each of the nine subjects showed that the positive after image disappeared sooner (on average) when the disc was moving than when it was stationary. The average time of disappearance for the five trials for the nine subjects was 2·1 sec. from the instant of appearance of the after image when there was movement, and 4·2 sec. when there was no movement. There were 41 cases when in the trial with movement the after image disappeared sooner than in the

following trial without movement, and only four cases with the opposite result. These results are clearly significant.

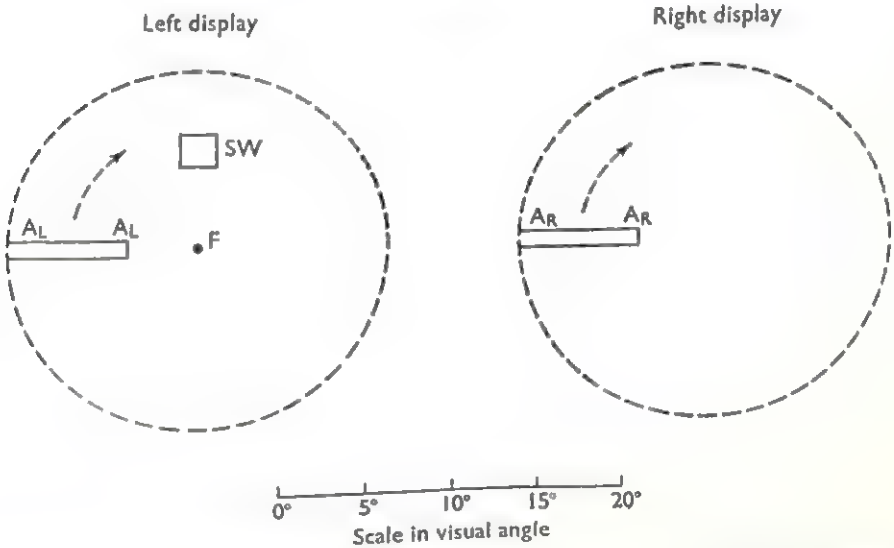
We attempted to do a somewhat similar experiment with negative after images, by altering the apparatus so that after the exposure of the square the subject's left eye now saw a white field. We did not find conditions under which the negative after image was sufficiently stable to allow satisfactory quantitative results. Nevertheless most subjects felt fairly sure that on those occasions when the after image *was* relatively stable, movement masking did occur.

Experiment 3

Monocular and binocular masking

MacKay (1960) has reported a similar phenomenon to what we have called Binocular Movement Masking when the moving object is seen by the *same* eye as the static object. A summary of his results is given in our earlier paper (p. 107). We therefore did an experiment to compare the monocular and binocular effects.

FIGURE 4



The arm A_L was only seen by the subject in the monocular trials, and the arm A_R was only seen in the binocular trials. The square SW and the fixation point F were seen in all trials. The dotted circles, i.e. the boundaries of the discs, were hardly visible against the black background.

Apparatus

The apparatus was slightly modified from that used in Experiment 2 so as to produce the displays shown in Figure 4. The display to the right eye was unaltered, except that the illumination was reduced so that the luminance of the arm (A_RA_R in Fig. 4) was 1·0 ft. L. In the left chamber (Fig. 1) a half aluminized mirror was introduced in a position symmetrical with that of the fully reflecting mirror MM, and a second disc (similar to that seen by the right eye) was also introduced. *Through* this half aluminized mirror the subject saw with his left eye the fixation spot F and square window SW, and *reflected* from it he saw the white arm A_LA_L. The luminances of the square window SW, the fixation spot F, and the arm A_LA_L were all adjusted to 1·0 ft. L. (i.e. to the same value as that of the arm A_RA_R). The motor driving the two discs worked at the same speed (12 r.p.m.), so that when the arms started from the position indicated in Figure 4 they would cross the middle of the square 1·25 sec. later.

Procedure

In tests of monocular masking the left disc was illuminated with the arm A_LA_L in its starting position (as shown in Fig. 4), and the fixation mark F was switched on. The subject was told to fixate, and 2 sec. later the square window SW was switched on and the disc began to rotate. Any disappearance of the square window was recorded by the subject on the recorder. After one revolution (i.e. 5 sec.) the disc was stopped and the trial ended. The right hand disc was not illuminated and the arm A_RA_R was invisible to the subject. In tests of binocular masking the procedure was exactly the same, except that in these trials the right hand disc was used, and the arm A_LA_L was invisible.

There were 12 subjects, and after a few practice trials, each was given five trials under each condition. Half the subjects were given the monocular trials first, and the other half the binocular trials.

Results

The results showed quite clear evidence of movement masking under both conditions; and the qualitative descriptions of the phenomenon were very similar to those reported in our earlier paper. In the present experiment masking was reported more frequently under binocular conditions than under monocular conditions. In the binocular trials there were 50 out of a possible 60 disappearances, but in the monocular trials there were only 29 out of a possible 60. Of the 12 individual subjects nine showed more disappearances under the binocular conditions, one showed the reverse, and two showed equal numbers of disappearances. The vast majority of the disappearances were reported as occurring very shortly after the arm crossed the square window.

Discussion

The results of Experiment 1 were surprising in that they showed a "negative" result of attention, i.e. that attention to an object made it more liable to movement masking, rather than more resistant to it. But it must be noted that this is not the only case in which attention has been reported to act in this direction. Thus Babington Smith (1961*a*,*b*) reported an experiment in which prolonged attention to one of a number of static objects in the periphery of the visual field caused that one to disappear before the others. We have confirmed Babington Smith's result in several further experiments, and we hope in a later paper to compare the conditions under which attention has positive and negative effects in peripheral vision.

The result of Experiment 2 showing the occurrence of movement masking with positive after images, and suggesting the same for negative after images, are hardly surprising on any theory of after images. It should, however, be pointed out that some of our subjects had seen movement masking with real objects, and were therefore likely to be expecting the same effect with after images. Inspection of the results does not suggest that their previous experience was important.

In Experiment 3 the important finding was that movement masking can occur both monocularly and binocularly. The occurrence of monocular movement masking confirms MacKay's findings (but with a display very different from his). The results of Experiment 3 suggest that with this display the tendency to binocular masking is much stronger than that towards monocular masking, but the evidence, of course, would not justify a generalization to other experimental conditions.

We should like to express our thanks to Professor Zangwill for encouragement, and to the Medical Research Council for a personal grant to one of us (V.T.) and for an apparatus grant.

References

Averbach, E., and Sperling, G. S. (1960). Short term storage of information in vision. *4th London Symposium on Information Theory*, page 196.

Babington Smith, B. (1961*a*). An unexpected effect of attention in peripheral vision. *Nature*, **189**, 776.

Babington Smith, B. (1961*b*). Effect of attention in peripheral vision. *Nature*, **191**, 732.

Brindley, G. S. (1962). Two new properties of foveal after-images and a photochemical hypothesis to explain them. *J. Physiol.*, **164**, 168–79.

Grindley, G. C. (1931). Psychological factors in peripheral vision. *Med. Res. Council Spec. Rep. Series*, No. 163, London.

Grindley, G. C., and Townsend, V. (1965). Binocular masking induced by a moving object. *Quart. J. exp. Psychol.*, **17**, 97–109.

Helmholtz, H. von (1925). *Handbuch der Physiologischen Optik*. English Translation, Southall, J. P. C., Vol. 3, Chap. 31, page 455.

MacKay, D. M. (1960). Monocular "rivalry" between stabilised and unstabilised retinal images. *Nature*, **185**, 834.

Manuscript received 22nd June, 1966.

TRANSFER OF TRAINING AFTER GUIDANCE OR PRACTICE

BY

A. W. MACRAE* and D. H. HOLDING

From the University of Leeds

On a pursuit tracking apparatus presenting target courses of three levels of complexity, provision was made for either normal practice or forced-response guidance; the guidance training was gained by holding the control knob during automatic tracking. After five training trials on the most complex course, or on the simplest course, subjects were transferred to the intermediate course. The effects of guidance on the intermediate course were also examined.

Normal practice on the simple course produced more transfer than normal practice on the most complex. Further, guidance on the complex course gave significantly better transfer than did practice on that course. The superiority of guidance is tentatively ascribed to the opportunity it provides for the development of anticipation.

Introduction

Physical guidance methods are those in which learning is achieved by the external manipulation of the learner's movements. Studies of the advantages and disadvantages of these methods are gradually accumulating; most of the available evidence has been reviewed by Holding (1965). Work has mainly been confined to the effects of guidance on subsequent performance of the guided task, there having been little concern with the results of transfer to a different task.

One of the early studies of guidance (Ludgate, 1924) deals with the transfer of learning between easy and difficult stylus mazes. To administer guidance, the experimenter grasped the stylus just below the subject's hand, dragging it manually along the correct path. The group guided in this way on a first maze transferred their training more readily to the second maze than subjects who had practised normally on the first, an effect which increased according to the number of errors averted by the guidance procedure.

Maze-learning is a form of procedural task, in which the element of motor adjustment is at a minimum. At the opposite pole are continuous adjustive skills like pursuit tracking, for which the transfer of brief guidance was investigated by Holding (1959). Guidance appeared to transfer somewhat better than normal practice on this kind of task, although the effect did not reach significance and was a subsidiary outcome of the experiment. The present study is designed to examine more closely the transfer effects arising from guidance training in a pursuit task.

Method

Subjects

Tracking scores were obtained from 60 male students, undergraduate and postgraduate, from various faculties.

Apparatus

The equipment comprised a target course generator, display and control units, a servo guidance system, error processing and recording devices, and ancillary power supply units.

Target courses were provided by a motor and multiple gear box, which generated and mixed any combination of one to four sine waves. Resultant motions were

* Now at University College of North Wales, Bangor.

transduced by a two-phase resolver for electrical transmission to the guidance and display units.

The display consisted of a defocussed green spot (approx. 3 mm. diameter) moving over a maximum 90 degrees of an arc of radius 7 cm. on an oscilloscope tube. This was tracked by a perspex pointer of length equal to the target motion radius, illuminated at the tip by an amber neon. The pointer was operated by a serrated, 4 cm. control knob situated 10 cm. below and 3 cm. to the right of the display centre, having a gear ratio of 1 : 1 with respect to the pointer.

Under conditions of normal practice the subject moved the control knob, but when guidance was to be given he grasped the knob lightly in such a way as to permit passive movement over the necessary range. Automatic tracking movements of the control were activated by a servo motor and torque amplifier linked to the target motion. The output of all servo systems represents a compromise between lag and damping; in this case values were chosen which minimized the error score as far as was compatible with the elimination of hunting. The resulting performance resembled that of a human subject in eliminating any rapid oscillation at the cost of some lag, the root mean square (RMS) error under automatic operation averaging 3 degrees, or half the value of the error shown by practised human subjects on the easiest course.

Difference voltages between the display and control indicators were electronically squared and integrated to provide the basis of the RMS error score, square roots being derived by the experimenter from the scores accumulated by a dekatron counter.

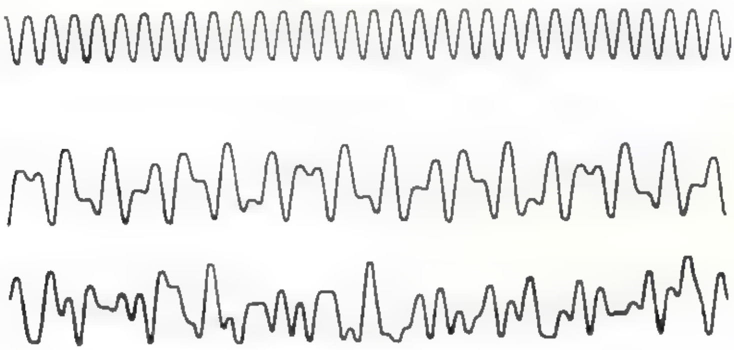
Design

Three target courses were used, all at an RMS amplitude of 22 degrees (approx. 2·7 cm.):

Simple: a single sine wave at 0·7 cycles per sec. (cps.).
Intermediate: summation of two sine waves at 0·4 cps. and 0·7 cps.
Complex: summation of four sine waves at 0·3, 0·4, 0·7 and 1·0 cps.

Figure 1 gives an indication of the appearance of the three courses over time.

FIGURE 1



Examples of the three target courses.

Subjects were allocated in order of appearance to six experimental groups, the sequence of conditions being changed daily. All subjects performed five ½-min. trials, interspersed with 1-min. rest pauses, on one of the forms of prior training shown below:

Group	*Prior training*
SP	practice on simple course
SG	guidance on simple course
IP	practice on intermediate course
IG	guidance on intermediate course
CP	practice on complex course
CG	guidance on complex course

After the prior training and a 3-min. break, subjects in all groups then performed five trials of normal practice on the intermediate course. Group IP thus formed a control group against whose performance the transfer scores of the other groups were compared.

The instructions which were given related the method of scoring to the aims of tracking, and emphasized that subjects would later be transferred to different target courses. Knowledge of results, in the form of dekatron scores, was given after each scored trial.

Results

The amount learned during normal practice tended to decrease with increasing course complexity. This effect is shown in Table I, which lists the error scores at the beginning and end of normal practice during the first half of the experiment.

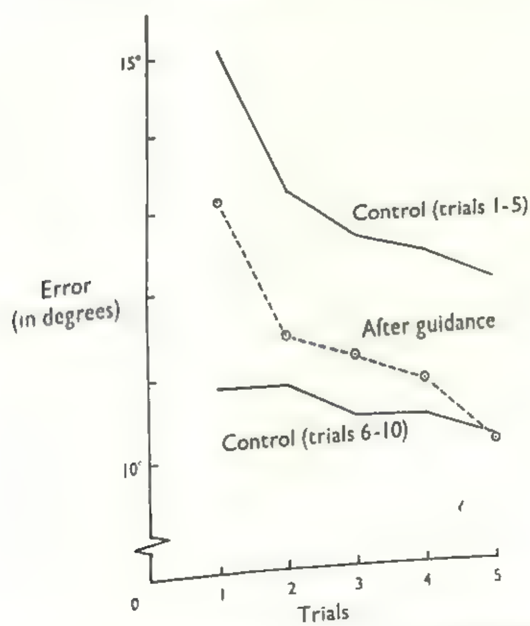
TABLE I

Performance During Prior Practice

Course	*R.M.S. error (degrees)*		*Percentage improvement*
	Trial 1	*Trial 5*	
Simple ..	8·9	6·0	33
Intermediate ..	15·1	12·0	21
Complex ..	17·3	15·4	11

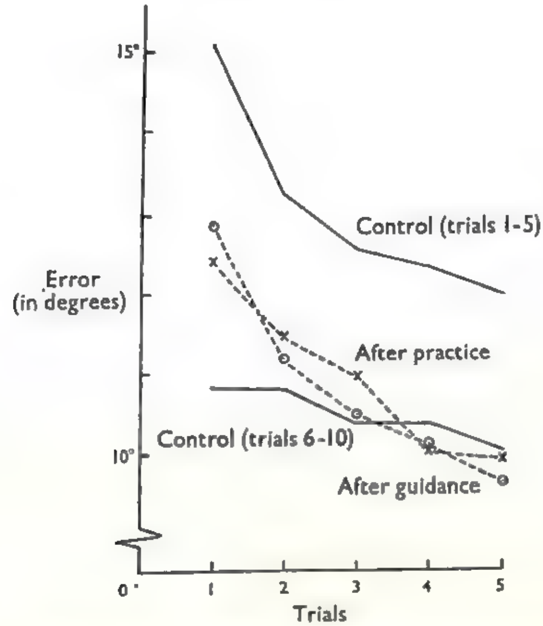
The transfer scores obtained during the second half of the experiment are presented separately. Figure 2 shows the effects of prior guidance (IG) on the intermediate course, while Figures 3 and 4 show the learning curves for groups transferred to IP from the simple and from the complex courses. The IP control curves are reproduced in each case, with the second half of practice shown as a separate curve.

FIGURE 2



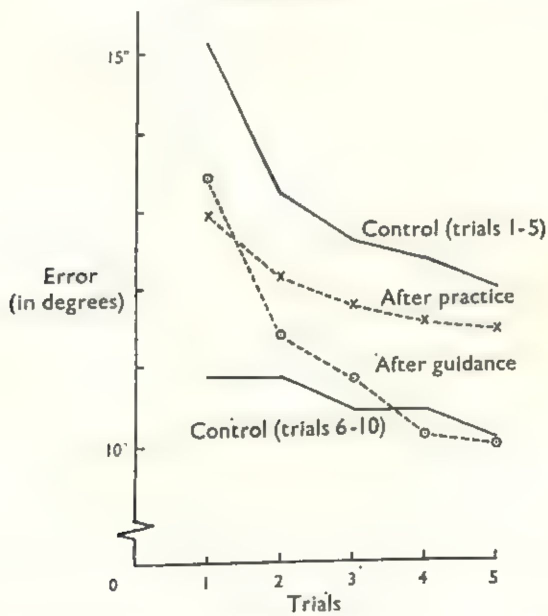
The results of prior guidance on the intermediate course.

FIGURE 3



The transfer effects of prior practice or guidance on the simple course.

FIGURE 4



The transfer effects of prior practice or guidance on the complex course.

The first transfer trials

Inspection of the graphs shows that performance on the first trial of the transfer task is not representative of later performance, since the guided subjects improve more rapidly than do the prior practice groups; it has been noted in other types of

task (Holding and Macrae, 1966) that the first trial of practice after guidance may be unduly depressed. However, since some interest conventionally attaches to the first trial of a transfer task (Gagné, Foster and Crowley, 1948) an analysis of variance was performed to compare the first IP trial with the first trials of all transferred groups on intermediate practice. F is 3·52, for 5 and 54 *d.f.*, indicating a significant ($p < 0{\cdot}01$) overall difference between groups. Application of Duncan's (1955) range test does not separate the transfer groups, but distinguishes ($p < 0{\cdot}05$) all of these from group IP. It may therefore be inferred that significant initial transfer arose from all forms of pre-training.

Final trials: guidance for the same course

Further analyses were based upon the final performance of all groups. To assess the efficiency of guidance on the intermediate course as training for that course, a comparison was made between the end points of Figure 2; i.e. group IP trial 5, group IG transfer trial 5, and group IP trial 10. This procedure involves allotting separate degrees of freedom to the IP subjects at the end of the first and second halves of training, at the expense of some contamination of the comparison. Analysis of variance shows F to be 3·58, for 2 and 27 *d.f.*, so that the overall difference appears significant ($p < 0{\cdot}05$). Duncan's range test separates IP/5 from IG/5 and IP/10 together, at the same level of significance. The implication is that five trials of guidance followed by five trials of normal practice give the same amount of learning as 10 trials of normal practice, the level reached being significantly better than performance after five trials of normal practice.

Final trials: transfer of guidance and practice

Although Figures 3 and 4 show separately the results of transfer from the simple and complex courses, it is convenient to combine them for analysis. The final error scores reached by the four transferred groups are as follows:

	SG	*SP*	*CG*	*CP*
RMS Error . .	9·6°	9·9°	10·0°	11·5°

An overall analysis is significant ($p < 0{\cdot}05$; F is 3·17, 3 and 36 *d.f.*). Duncan's range test separates group CP from all the other three groups, which do not themselves differ significantly. It appears therefore that practice on the simple course (SP) gives better ($p < 0{\cdot}05$) final results than practice on the complex course (CP). Further, pre-training on the complex course is better (p) 0·05) effected by guidance (CG) than by normal practice (CP). With pre-training on the simple course the results are slightly better after guidance (SG), but the difference is not significant.

Discussion

The pattern which emerges from the results is one of substantial transfer by all groups; initial transfer, assessed by representing the difference between the first control score and the first transfer score as a percentage of the total improvement shown by the control group, ranges from 33 to 53 per cent. Following the first transfer trial all but one of the transferred groups make faster progress than the control group, finishing with scores comparable to the tenth and final trial of normal practice, and thus with 100 per cent. transfer. The exception is the CP group, whose members received pre-training in the form of normal practice on the complex task; the transfer scores of this group improve less rapidly than those of the other groups,

their final error score remaining well above the performance of the control group and of the corresponding guided group. The complex practice subjects were in no way anomalous, so that the superior transfer value of the guided training therefore requires explanation.

The possible reverse argument, that complex practice is in some way detrimental by comparison with the other treatments, appears improbable. It seems perverse to suggest that normal practice is an abnormal condition while, in any case, the argument would merely rest upon the supposed incongruity of a transfer group whose scores fail to converge with those of others in the later stages of a second task. However, the degree of learning finally reached may often depend upon the form taken by initial training. In the transfer of tracking skill Gordon (1959), for instance, shows many transfer curves which do not converge over nine trials and Holding (1962) reports differences which persist throughout a week of transfer performance. The view is therefore adopted that practice on the simple task is a more appropriate preparation for performance on the transfer course than is practice on the complex task and that, for different reasons, complex guidance is more effective than complex practice.

The salient contribution of guidance in this context lies in accustoming the subject to a low error rate while relieving the task of motor demands, so that his commitment becomes largely perceptual. For similar reasons it appears (Holding and Macrae, 1966) that the efficiency of guidance rises as the need for perceptual-motor translation increases the involvement of central processes. The guidance technique supplies some *response information* but acts to advantage in situations requiring a high proportion of *task information*, which must include both immediate and anticipated features of the perceptual display and information concerning the relationship between display and control characteristics.

The acquisition of task information, at least on the complex course, must presumably have been facilitated by the guidance method to an extent precluded by the motor demands of normal practice. It is possible that response information is also involved, in that complex guidance may bring home to the subject what level of activity is needed. However this explanation does not hold for prior guidance on the simple course, so that it seems more important to notice that the guidance procedure allows the subject to "filter out" the response information when necessary. He is thus free to turn his attention to the task information as necessary, in contrast to the subject practising normally whose neglect of response is immediately followed by a disrupting change in the perceptual display.

Guidance makes the subject's task "easier" and, since normal practice on the simple course produced more transfer than normal practice on the complex version, there is an apparent superiority of easy-to-difficult over difficult-to-easy transfer. However, as Holding (1962) has shown, such effects depend upon the balance between learning more, where difficult tasks include the essential components of the easier ones, as opposed to learning better when practice on the easier tasks makes for performance habits which transfer well to more difficult versions. In the present experiment practising the simpler tasks seems likely to encourage anticipation of the target course, which will produce a high level of accuracy on that course and the means of maintaining it on the intermediate course.

The amount of learning during normal practice on the simpler task is quite substantial, as shown in Table I. In contrast to this, and to the stable error scores of the automatic tracking device, human tracking of the complex course began at an error level twice as high and showed only one-third of the improvement. Subjects practising on the complex course had therefore little learning to transfer, whereas

subjects on the simple course made gains which evidently were not specific to that course. As Poulton (1956) has emphasized, a great deal of the improvement in skilled performance depends upon obviating by forward prediction what would otherwise constitute a decision time delay; it seems probable that a tendency to anticipate will give subjects an advantage on transfer to a new course.

Thus, although subjects receiving guidance on the simple or complex courses do not meet identical course characteristics on transfer to the intermediate task, it seems reasonable to assume that their performance is biassed towards perceptual anticipation and away from mere following of the target course. The kind of task information they have acquired is therefore favourable to transfer. Further, any tendency to anticipate would not be reflected in high performance immediately on transfer, but might well result in the observed steady improvement throughout the transfer trials.

This experiment forms part of a project supported by the Science Research Council.

References

DUNCAN, D. B. (1955). Multiple range and multiple F tests. *Biometrics*, **11**, 1–42.

GAGNÉ, R. M., FOSTER, H., and CROWLEY, M. E. (1948). The measurement of transfer of training. *Psychol. Bull.*, **45**, 97–130.

GORDON, N. B. (1959). Learning a motor task under varied display conditions. *J. exp. Psychol.*, **57**, 65–73.

HOLDING, D. H. (1959). Guidance in pursuit tracking. *J. exp. Psychol.*, **57**, 362–6.

HOLDING, D. H. (1962). Transfer between difficult and easy tasks. *Brit. J. Psychol.*, **53**, 397–407.

HOLDING, D. H. (1965). *Principles of Training.* London: Pergamon.

HOLDING, D. H., and MACRAE, A. W. (1966). Rate and force of guidance in perceptual-motor tasks with reversed or random spatial correspondence. *Ergonomics*, **9**, 289–96.

LUDGATE, K. E. (1924). The effect of manual guidance upon maze learning. *Psychol. Monogr.*, 33, No. 148.

POULTON, E. C. (1956). On prediction in skilled movement. *Psychol. Bull.*, **54**, 407–78.

Manuscript received 17th February, 1966.

TIMES FOR TRANSITIONS BETWEEN HAND AND FOOT RESPONSES IN A SELF-PACED TASK

BY

PATRICK M. A. RABBITT

From the Medical Research Council, Applied Psychology Research Unit, Cambridge

In a self-paced task subjects responded to each of four equally probable signals with a different one of their four limbs. Response times were examined as a function of the 16 possible transitions between limbs. Repeated responses were shown to be faster than any other transitions, while responses following responses with an ipsilateral limb were relatively slow. The implications of these results for models for the "repetition effect" are discussed.

INTRODUCTION

A response is made faster when it is repeated than when it follows some other response. This has been described as the "response repetition effect" (Bertelson, 1963, 1965). Under certain conditions of S–R mapping a response is also made more quickly if it follows another response made with the same hand than if it follows a response made with the contralateral hand (Rabbitt, 1965). These experiments show that a "repetition effect" may neither require the repetition of a particular stimulus nor the repetition of a particular response. This suggests that repetition effects may in some cases require the repetition of only part of a series of decisions which we must assume to be made in the C.N.S. in order to programme a motor act. In these terms it is logical to consider what characteristics of successive movements are critical in order for such facilitation to occur. The possibilities are limited by human anatomy; successive responses may be made with the same limb, with ipsilateral limbs of different kinds (right or left hand and foot), with contralateral limbs of the same kind (both hands or both legs) and with diagonally-opposite limbs (i.e. hand and foot on opposite sides).

While Blyth (1962, 1963, 1964) did not directly examine transition effects, his findings provide the only relevant data known to the writer:

If successive responses (R_1 and R_2) are made with different limbs to two stimuli (S_1 and S_2) which follow each other at a very brief interval, R_2 takes longer than R_1. The relative delay of R_2 has been taken as evidence for a "psychological refractory period" (Vince, 1948; Welford, A. T., 1952). Blyth (1962) examined variations in the delay of R_2 between the 12 possible cases in which R_1 and R_2 are made with different limbs. He found that R_2 was faster when the two responses required successive movements with ipsilateral limbs, or with contralateral limbs of the same kind, than when they required successive movements with diagonally-opposite limbs.

It cannot be argued from this result that transitions between limbs on the same side, or between limbs of the same kind, are always faster than transitions between different kinds of limbs on opposite sides of the body. In Blyth's experiment S_2 occurred either 100 millisec. or 200 millisec. after S_1, whereas his subjects took from 350 to 450 millisec. to make R_1. His results therefore only cover the special case where S_2 occurs before R_1 is completed. Moreover, some of Blyth's (1962) other experimental results are, at first sight, difficult to reconcile with the hypothesis of fast ipsilateral transition. For example, he found that when subjects responded to each of two equi-probable signals in a positional display with one of two different

limbs, response latencies for any limb were shorter when the alternative response was made with a diagonally-opposite limb than when it was made with an ipsilateral limb (Blyth, 1963). Consistent with this latter result was his finding that, in a four-choice/four-limb response task with a positional display, errors are more commonly confusions between ipsilateral limbs than confusions between contralateral or diagonally-opposite limbs.

Blyth's extremely thorough study therefore offers the different suggestions that transitions between ipsilateral limbs are made relatively quickly in some situations, while selection between ipsilateral limbs is relatively slow and inaccurate in others. An experiment was made to directly examine transition effects in a self-paced task in which subjects responded to each of four equi-probable signals with a different one of their four limbs.

Method

Procedure. Each subject sat at a desk on which an In-line numeral display unit faced him. Using a Creed tape-reader, any one of the digits, 1, 2, 3 or 4 could be programmed to appear 1 in. high in illuminated outline against the 1·8 in × 1·3 in. background of the display. Subjects responded to each of these digits by pressing down a different one of four switches. The display changed to a new signal within 100 millisec. of the switch being again opened. A brief blink of the display indicated a change when the same digit was repeated.

Of the 16 different mappings of switches on digits 12 were allocated at random, one to each of the subjects used in the experiment. Two hand-switches were telegraph keys, adjusted so that all subjects could comfortably grasp them while leaning their fore-arms on the desk. Two micro-switches were mounted under spring-loaded foot-pedals on an adjustable frame beneath the desk. All subjects could comfortably rest their feet on the pedals, and only a slight additional pressure was necessary for them to close either switch. Adjustments for each subject ensured that his balance was not disturbed by the movement necessary to close any switch. Subjects wore their shoes.

Fifteen different runs of 97 signals were programmed—five runs on each of three different programme tapes. Within each run digits appeared in random order, with the constraint that each of the 16 possible transitions between digits occurred six times. A version of SETAR (Welford, N., 1952) recorded on punched tape each stimulus presented, the response the subject made and the elapsed time (within 0·01 sec.) since the termination of his last response (i.e. reaction time + a constant of 100 millisec. required for the display change).

The apparatus gave the subject no immediate error-indication, but an error-counter provided him with a cumulative score after each run. He was however, given a mean response-time for each run, obtained with a stopwatch. Fast and accurate performance was urged and commended.

Each subject was tested once, for 50 min. Testing took place in a dimly lit, sound and vibration-proof cubical. The experimenter first explained the task, and ran a practice trial of 20 stimuli and responses. Subjects were then run through the three programme tapes. There was a 5–10 min. pause after each tape. Two of the 12 subjects received the three tapes in each of the six possible orders.

Subjects. These were 12 men aged 21 to 35 years drawn from the A.P.R.U. panel of paid civilian volunteers. All had normal vision and motor control.

Results

A print-out of SETAR output was analysed by eye to count the number and kinds of errors committed, and to calculate mean reaction-times for each of the 16 possible transitions between limb-movements.

Errors. 4·8 per cent. of all responses were wrong. Because of the small numbers involved comparisons of the incidence of the 12 possible erroneous substitutions were not statistically significant. Errors were therefore separated into three classes depending on whether they represented substitutions by ipsilateral, contralateral or diagonally-opposite limbs. These data are shown in Table I.

TABLE I

BREAKDOWN OF ERRORS INTO CLASSES OF SUBSTITUTIONS BETWEEN LIMBS

Substitution by ipsilateral limb	55·0 per cent.
Substitution by contralateral limb	25·2 per cent.
Substitution by diagonally opposite limb ..	19·8 per cent.
Total number of errors	838

Response times. Mean response times for each of the 16 possible correct transitions between limbs are set out in Table II. Analysis of variance gave significant terms for differences between subjects ($p < 0{\cdot}01$) and between transition-classes ($p < 0{\cdot}001$). The error-term from this analysis was used to calculate S^2 (Studentized range statistic, cf. Winer, 1962, p. 77).

TABLE II

RESPONSE TIMES (IN MILLISEC.) FOR 16 POSSIBLE TRANSITIONS BETWEEN LIMBS IN A FOUR-CHOICE/FOUR-LIMB SELF-PACED TASK

	Preceding response made with			
	Right hand	*Left hand*	*Right foot*	*Left foot*
Responding limb				
Right hand ..	568 ($\sigma = 65{\cdot}0$)	642 ($\sigma = 87{\cdot}8$)	732 ($\sigma = 94{\cdot}0$)	687 ($\sigma = 76{\cdot}9$)
Left hand ..	648 ($\sigma = 52{\cdot}6$)	560 ($\sigma = 51{\cdot}4$)	675 ($\sigma = 59{\cdot}0$)	683 ($\sigma = 63{\cdot}9$)
Right foot ..	735 ($\sigma = 63{\cdot}9$)	705 ($\sigma = 48{\cdot}2$)	600 ($\sigma = 56{\cdot}1$)	711 ($\sigma = 61{\cdot}5$)
Left foot ..	769 ($\sigma = 48{\cdot}4$)	792 ($\sigma = 70{\cdot}4$)	732 ($\sigma = 90{\cdot}1$)	604 ($\sigma = 50{\cdot}1$)

Four transition-classes may be considered for each limb: repeats, contralateral, ipsilateral and diagonal. S^2 was used to calculate the significance of differences between the total for these categories:

Right hand. Repeated responses were significantly faster than contralateral transitions ($p < 0{\cdot}05$), than diagonal transitions ($p < 0{\cdot}01$) or than ipsilateral transitions ($p < 0{\cdot}01$). Contralateral transitions were faster than ipsilateral transitions ($p < 0{\cdot}01$). No other differences were significant.

Left hand, left foot, right foot. These limbs showed patterns of results similar to each other. That is, repeats were faster than any other transition-class (p) 0·01). but no other differences were significant. To determine whether any trends were significant the data for all 16 possible transitions were collapsed over these four transition-classes. Means for these classes are given in Table III. By re-calculated S^2 only the following differences were significant: repeats are faster than all other responses ($p < 0{\cdot}01$); contralateral transitions are faster than ipsilateral transitions ($p < 0{\cdot}01$); diagonal transitions are faster than ipsilateral transitions ($p < 0{\cdot}05$).

The data for 16 transition classes can be collapsed in a similar way to allow comparisons between all possible arbitrary combinations of four sets. Nearly all of these possible segmentations of the data are obviously meaningless in terms of the logic of the experiment, but certain further comparisons were necessary.

TABLE III

OVERALL MEAN RESPONSE TIMES (IN MILLISEC.) FOR FOUR CLASSES OF TRANSITIONS BETWEEN LIMB MOVEMENTS IN A SELF-PACED CHOICE-RESPONSE TASK

Transition category	$\bar{x}$	σ
Repeats	583	49·6
Contralateral ..	683	83·9
Diagonal	709	69·7
Ipsilateral ..	744	61·4

(1) In order to test whether movements with any one particular limb relatively facilitated or inhibited all subsequent responses, the 16 transition-classes were collapsed so as to separate out four classes of responses with each of the four limbs. No significant differences were observed, indicating that transition effects are here specific to *relationships* between limbs rather than specific to *limbs*.

(2) The data were similarly collapsed to compare responses made with each of the four limbs (means for these classes are right hand, 657 millisec.; left hand, 641 millisec.; right foot, 670 millisec.; left foot, 725 millisec.). Response times for the hands are not significantly different ($p > 0·05$) but response times for either hand are faster than for either foot ($p < 0·05$). Response times for the right foot are faster than for the left foot ($p < 0·01$). The 16 possible transitions were collapsed into two classes, and three of the possible resulting breakdowns were selected for comparison by recalculated S^2.

(1) The class of responses involving successive movements with limbs on the same side of the body was not significantly different from the class of responses involving successive movements with limbs on opposite sides of the body.

(2) There was no difference between responses following responses made with hands and responses following responses made with the feet.

(3) Responses following responses made with a limb of the same type (repeats + contralateral transitions) were faster ($p < 0·01$) than responses following responses made with a limb of a different type (i.e. ipsilateral + diagonal transitions). (This difference can be mainly attributed to the fact that repeated responses are much faster than any others.)

Effects of practice. To determine whether differences in response-times between transition classes were affected by practice, mean response-times for repeats, contralateral transitions, diagonal transitions and ipsilateral transitions were calculated separately for each of the three successive tapes of five runs (i.e. of 485 responses). An analysis of variance on these data gave significant terms for differences between transition-classes ($p < 0·01$) and for improvement with practice ($p < 0·01$). There was no significant interaction between transitions and practice, so that the present data allow no comment on the role of practice in relation to this effect.

DISCUSSION

Response-time varies with anatomical relationships between limbs with which successive movements are made. There is no suggestion that movements with any

one limb, with limbs of either type (i.e. hand or foot), or with limbs on either side relatively facilitate or inhibit all subsequent responses. The effects observed may be summarized as follows:

(1) Repeated responses are faster than any other class of transitions.

(2) Contralateral transitions are faster than ipsilateral transitions.

(3) Diagonal transitions may be faster than ipsilateral transitions (this difference barely attains statistical significance at $p < 0{\cdot}05$).

The first of these points simply confirms the existence of a "response repetition effect" in a new context.

The rank-order of RT's in other transitions-classes (points 2 and 3) parallels Blyth's finding in some of his experiments that subjects select more quickly between contralateral limbs or between diagonally-opposite limbs than between ipsilateral limbs (Blyth 1962, 1963). In the present experiment, as in all of Blyth's, confusions/substititions between ipsilateral limbs were more frequent than between contralatera' or diagonally opposite limbs. On the other hand, the present results contrast with Blyth's finding that the "psychological refractory" delay is longer for diagonal transitions than for ipsilateral transitions.

Blyth found that the largest class of errors were confusions between ipsilateral limbs (40–90 per cent. depending on the type of display used). In the present experiment 55 per cent. of errors were also of this type. Blyth showed by experiment that confusion errors occurring because of simultaneous response by two limbs were most frequent with ipsilateral responses. It is reasonable to suppose that subjects became aware of this tendency, and might try to reduce errors by inhibiting an ipsilateral limb-movement when making any response. This would increase the time taken to choose between ipsilateral limbs or to make ipsilateral transitions. In contrast, when subjects know in advance that successive movements are to be made with ipsilateral limbs it would be advantageous *not* to suppress this tendency. However, this strategy would have the liability that subjects would tend to make R_1 and R_2 simultaneously, or even to reverse the order of responses. Blyth's analysis of errors in his refractoriness experiment (Blyth, 1962, p. 180–181) does indeed suggest that reversals are more common when R_1 and R_2 are made with ipsilateral limbs. This trend is particularly marked where R_1 should have been made with a foot— so that it is possible to argue that a fast response with an ipsilateral hand has been made before the slower foot-response could "catch up."

This simple hypothesis might account for Blyth's finding that ipsilateral transitions are faster than diagonally-opposite transitions, but it does not explain why he found that contralateral transitions were also relatively fast. The assumption that a response with any limb is accompanied by tendencies to make simultaneous responses with *both* an ipsilateral and a contralateral limb is not supported by the data (though weak trends in Blyth's error-analysis are consistent with this hypothesis). Among possible explanations we may suggest:

(1) That relatively fast contralateral transitions are due to the particular display-control relationship used in the refractoriness experiment (cf. Blyth, 1962).

(2) The differences observed may occur because, in everyday life, transitions between contralateral limbs are more practised than transitions between diagonally-opposite limbs.

Since S–R mappings were systematically varied between subjects it is probable that transition effects in the present experiment relate to difficulties in the programming of sequences of movements rather than to the peculiarities of any

particular S–R relationship. Moreover, the relatively slow transitions between ipsilateral limbs reflect a trend (i.e. slow ipsilateral choices) observed by Blyth in two different S–R mappings. Whether or not a hypothesis can be found to adequately describe the interactions of all possible variables, the present data confirm that some transition-effects relate to differences in the ease with which various sequences of movements may be programmed.

The writer is most grateful to Dr. J. A. Leonard for many discussions, and for the use of his apparatus.

References

Bertelson, P. (1963). S–R relationships and reaction-times to new versus repeated signals in a serial task. *J. exp. Psychol.*, **65**, 478–84.

Bertelson, P. (1965). Serial choice reaction-time as a function of response versus signal-and-response repetition. *Nature*, **206**, 217–8.

Blyth, K. W. (1962). *Experiments on Choice Reactions with the Hands and Feet.* Unpublished Ph.D. thesis, University of Cambridge.

Blyth, K. W. (1963). Ipsilateral confusion in 2-choice and 4-choice responses with the hands and feet. *Nature*, **199**, 1312.

Blyth, K. W. (1964). Errors in a further 4-choice reaction task with the hands and feet. *Nature*, **201**, 641.

Rabbitt, P. M. A. (1965). Response facilitation on repetition of a limb movement. *Brit. J. Psychol.*, **56**, 303–4.

Vince, M. (1948). The intermittency of control movements and the psychological refractory period. *Brit. J. Psychol.*, **38**, 149–57.

Welford, A. T. (1952). The psychological refractory period and the timing of high-speed performance—a review and a theory. *Brit. J. Psychol.*, **43**, 2–19.

Welford, N. (1952). An electronic digital recording machine—the SETAR. *J. Sci. Instrum.*, **29**, 1–4.

Winer, B. J. (1962). *Statistical Principles in Experimental Design*, New York: McGraw-Hill.

Manuscript received 10th June, 1966.

SPECIAL ARTICLE

THINGS, WORDS AND THE BRAIN*

BY

R. C. OLDFIELD

From the Institute of Experimental Psychology, University of Oxford

Our first duty, and special pleasure, this afternoon—and in this if in nothing else the lecturer and his audience will be at one—must be to offer our respects and congratulations to Sir Frederic Bartlett on being on the eve of completing the first 80 years of his life, and to express our admiration of the use to which he has put them. He was born seven years after the University of Cambridge had turned down a proposal to establish what might have been the first psychological laboratory in Europe. He started work soon afterwards I take it—for, like charity, psychology surely begins at home. Today, not only this Cambridge Psychological Laboratory but also those in the great majority of British universities, are bursting at the seams and turning students and researchers away from their doors. Sir Frederic cannot, of course, be held solely responsible for all this—progress; for some of it, perhaps, he might scarcely wish to be. But by the sustained quality of his own work, by his brilliant leadership at Cambridge and by the weight of his influence in the world outside—and especially in those quarters which may or may not choose to nourish fledgling disciplines—he stands alone among those who have contributed to the growth of our subject in this country.

For some of us here today—and some others who are not—no formal appreciation, however sincere and complete, could suffice. Things we would like to express are too personal to be appropriate on this occasion. But Sir Frederic will know that though unsaid, they are not unfelt.

One could wish—and not only because it would make the subject of my discourse more fitting—that Sir Frederic had more explicitly concerned himself with the psychology of language. Certainly I remember how much I got out of his penetrating remarks when long ago Gustav Stern's book *Meaning and Change of Meaning* was the subject of one of those early discussion classes—which otherwise provided undergraduates with opportunities for statistical research into the time-course of Bartlett's efforts to keep his pipe alight! Nor can we forget the part he played—gratefully acknowledged by Sir Henry Head—in the writing of *Aphasia and Kindred Disorders of Speech*, a book which was the early inspiration of more than one of us and which, for all the faults inseparable from pioneering, has still not had its day. At all events, I am going to talk about language—if only in a pretty restricted connection. We shall touch on aphasia, but I am afraid we cannot hope to get as far back as the meaning of meaning.

FINDING WORDS FOR SPEECH

When we speak or write we generate a sequence of words. In the present context we can take words for granted, but by no means all possible strings of them convey meaning. There are ways, however, in which the characteristics of permissible strings can to some degree be formally defined. And there are others by which

* The first Sir Frederic Bartlett Lecture, delivered in Cambridge on 12th July, 1966. Sir Frederic's eightieth birthday was on 20th October, 1966.

strings can be artificially generated on a stochastic basis so as to approximate, to a greater or less extent, and along a quantitative dimension, to meaningful passages. I am going to put myself at what many people may consider a fatal disadvantage this afternoon by neglecting these sequential relations between words in language as she is spoke. I do this so as to focus attention more closely on the problem of where each word comes from. Every speaker has at his disposal a vocabulary—a stock of words—from which he is able to produce items appropriate to the moment. As to the size of this stock, there have been various estimates. One I made some years back by a random sampling technique (Oldfield, 1963) suggested that, so far as *lexical units* are concerned, and disregarding *mere* compounds, the average, reasonably young, university-educated, kind of person knows the meaning of about 75,000 ± 3,500 words. The number ranged among individuals from some 55,000 to some 90,000. When he speaks, such a person will emit words in a way which at least from a purely linguistic point of view is tolerably meaningful at a rate of about two a sec. If we have no presuppositions as to how this is accomplished—and as scientists we ought not to have—it is a somewhat remarkable performance calling for investigation. For on the face of it, in a time which is on the average less than half a second, we are able to choose the appropriate item from among 75,000. Two of the first questions that suggest themselves are: "How is this 'stock' or 'store' organized, arranged and indexed?" and "By what means do we gain access to items in it?" These questions are clearly related. Organization implies relations—grouping and labelling of items. An access-system involves rules of procedure which operate in terms of the organization. In the absence of any organization the only rule is to examine each item in turn. With 75,000 words, and allowing $\frac{1}{2}$ sec. for each, this lecture—if I had not written it down beforehand—would come to an end somewhere about September, 1970, and I fear we could hardly wait to hear the end of it!

Looking around for some plausible form of economy in the storage and retrieval system, one thinks naturally of an ideal form which allows binary, dichotomous, search or some approximation to it. The items would each be labelled by a combination of characteristics in such a way that the presence or absence of one of them would enable half the items to be rejected by one decision, half the remainder by a second and so forth. On this basis we could sort $N=2^n$ items by n decisions—or conversely, we should need $\log_2 N$ decisions to sort N items. If N were $7{\cdot}5 \times 10^4$ the number of decisions needed, or bits of information to be processed would be about 16. Allowing $\frac{1}{2}$ sec. and as before for each item this would mean a rate of 32 bits per sec.—well within the limits experimentally estimated for various human skills including symbolic and linguistic ones.

But in such a system all items would be at the same distance—in terms of decision-steps—from the starting point, and all access-times would be equal. In practice this would be a disadvantage, because some words are needed more often than others. It might be better to arrange shorter access-times for these, even at the expense of longer ones for the rarer words. A good storage system might be one in which the mean access-time was minimized.

These are very general considerations, but they do suggest—so far as experiment is concerned—that two quantitative aspects of word-finding would repay attention, namely the times taken to find words and the commonness or rarity of the words in question. The latter conception is obviously the more difficult. But for present purposes we shall take as the measure of the commonness of a word the frequency with which it is found in a large textual sample. For the English language we have the Thorndike-Lorge (1944) Word Count for a set of samples totalling $4\frac{1}{2}$ million words of text. Many and serious criticisms can be levelled at both the construction,

and the use, of this. The odd thing is that in practice it doesn't work too badly. Reasonably clearcut relationships emerge in terms of the values it provides, and these latter are involved only in the logarithmic, not the absolute form.

There are basically two ways in which we can try to find out about the word-store and retrieval system. One is by studying word-production in a natural setting, as did Goldman-Eisler (1958) and others when they recorded the time-intervals between words in ordinary discourse and related them to the unexpectedness of the words, as estimated by the difficulty in replacing them correctly when deleted from the written record. Another way is by interrogating, so to speak—the storage and retrieval system by a simple task and a single stimulus. This avoids the complex contribution made by context, grammar and so forth, if at the expense of some artificiality. Both these approaches can be fruitful, but I shall confine attention this afternoon to the second. Various tasks and types of stimuli are available, and have been used experimentally. One example is the free-association experiment, in connection with which Thumb and Marbe as long ago as 1901 found that the time taken to respond to a single word-stimulus with the first word that came into the subject's head depended upon how common that word was among all the words given by all subjects to that stimulus. Such an experiment may bear several interpretations—not all of them throwing light on the storage system. Other examples are: reading a word, translating a word into another language and giving the name of a presented object. It is this latter case I want to consider now.

Naming objects

If we show someone an object, such as a pen, an anvil or a xylophone, and ask him to say what it is, the task is seemingly trivial, if for a healthy adult somewhat unaccustomed. He has first to identify it—and this means deciding that it is a member of a certain class—and then find the appropriate name. Nobody until J. McKeen Cattell in 1886 seems to have asked how long this process took, and after him, as happened with so many other things that worthy or his contemporaries did, nobody else bothered with it for the next 80 years or so, when Wingfield and I took it up again. When we started, Rochford and Williams (1962–5) were going, much more precisely than does the clinical neurologist, into object-naming by asphasics.

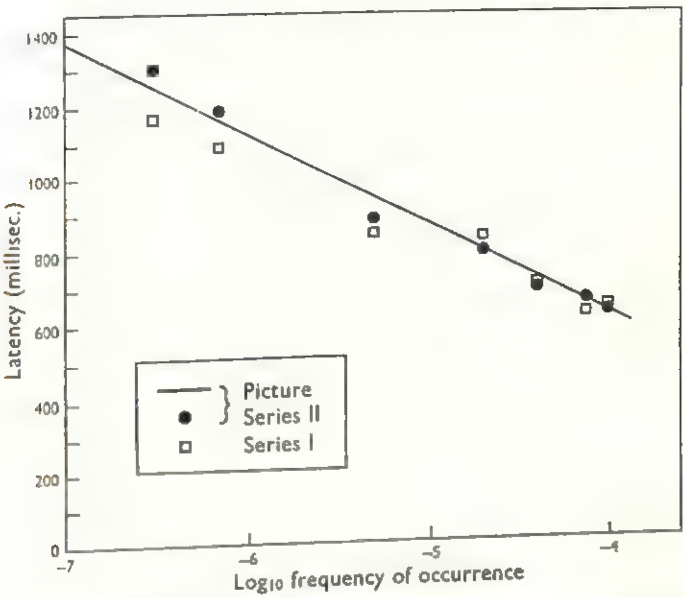
The chief results Wingfield and I (1965) obtained are shown in Figure 1. Different objects, in this case presented in the form of outline drawings, do take different times to name, and the times vary linearly with the logarithm of the frequency of their names as given by the Thorndike-Lorge Count. So far as these results are concerned, this is *not* due to the fact that rare words tend to be longer than common ones and might therefore require a longer time for the preparation of their articulatory patterns. Partial correlation analysis of the data in relation to word-length rules this out.

Word-frequency effects of this kind are, of course, not uncommon, though the time measured has more usually been the "visual duration threshold," or shortest time of presentation at a given intensity necessary for appropriate response. Howes and Solomon (1951), for instance, found the relationship in reading single words, while Wingfield (1966) demonstrated it for the pictures of objects we used in the naming experiment. Fraisse (1963) measured both visual duration thresholds and latencies for reading single French words and confirmed the logarithmic effect in both cases.

But we are perhaps a small step further forward. The mere existence of the effect rules out any simple binary organization of storage and retrieval. The opposite extreme, that of *seriatim* search, is equally excluded because in this case I would not

yet have got round to telling you so. We have to contemplate something in the nature of a process with two or more stages. Two immediate questions that arise about naming are: "What is the connection, in terms of *function*—or *mechanism*—between the *frequency* with which an object crops up (as reflected in the frequency with which its name occurs in texts) and the time it takes to name it?" Secondly, "How much of the time is taken up in perception or identification—and how much in the search for the word?"

FIGURE 1



Response-latency for naming as a function of name-frequency. (Reprinted from Oldfield and Wingfield, 1965, *Quart. J. exp. Psychol.*, **17**, 279.)

Probabilities, and *a fortiori*, frequencies, do *not* in themselves provide a basis for causal explanation. To relate the frequency more closely to the search mechanism we need a connection between frequency itself and the number of alternative objects, or names, which might have presented themselves with equal frequency, but didn't. For it is this number which decides how much search has to be undertaken to fix on the one correct item. We can make an estimate of this—after a fashion—in virtue of the relationship usually known as Zipf's Law (Zipf, 1935). This tells us the number of different words which all have a given frequency of occurrence. One form in which it may be stated is that the number N of different words found with frequency *f* in a large textual sample is proportional to *f* raised to a small negative power—which Zipf himself supposed to be −2.

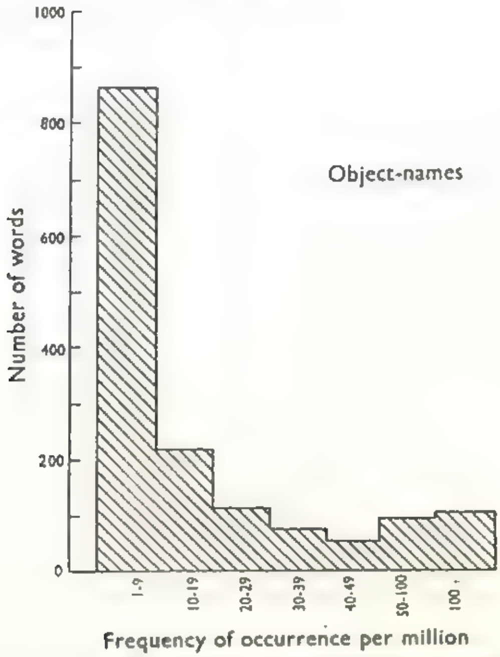
$$N = K.f^{-\alpha}$$

Zipf's Law, which is empirical, has given rise to much discussion and interpretation, about which I will only say that attempts, however valuable and interesting, such as that of Mandelbrot to derive it, or something similar, from more or less plausible and simple postulates do *not* render it trivial—as some have supposed in spite of Mandelbrot's (1965) own explicit disclaimer.

For the present we are concerned with the *names of objects* of the kind which are identifiable from their appearance, whether as pictures or real things, in isolation

from contexts or other objects. A rather tedious count of these I made from the Thorndike-Lorge list gave the results shown in Figure 2. The two most frequent

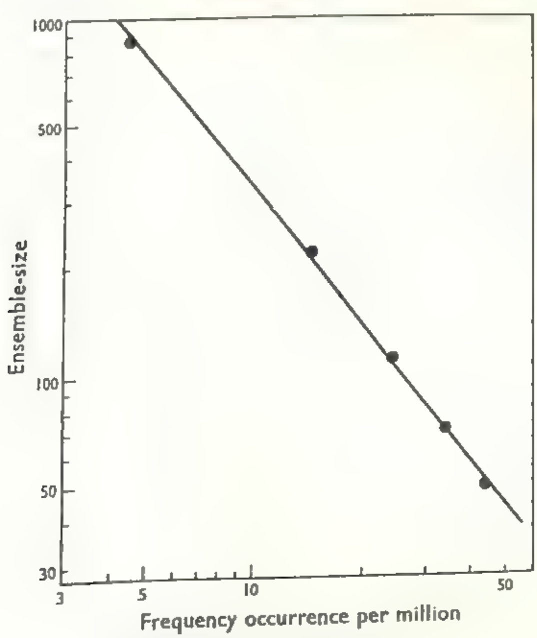
FIGURE 2



Frequency distribution of names in occurrence-frequency range shown.

classes are less deviant than they seem since their ranges are greater than the equal ones into which I have arbitrarily divided the rest. Apart from the two strays, the data fit a Zipf-type relationship very well (Fig. 3), the gradient, and hence α, the power of f, being $-1{\cdot}30$.

FIGURE 3



Zipf line for object-names.

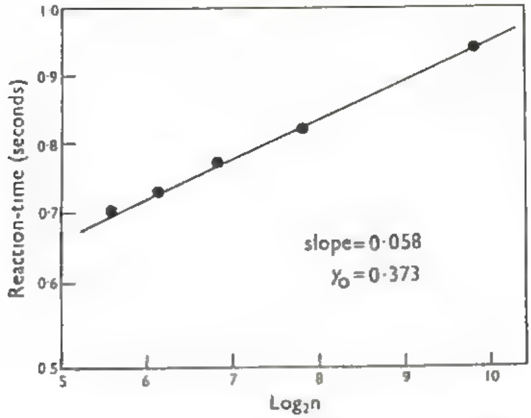
Treisman (1965), after an experiment in which she measured the (very much longer) latencies involved in finding a word to fill a gap in a verbal context, decided that a two-stage search model was more in keeping with her data than other possibilities she considered. The stages she proposed are not those I am going to suggest. In fact they are if anything the reverse, but her argument is too complex to do justice to at the moment. I will suggest, so far as naming is concerned, that the first stage consists in allotting the object to its correct frequency range, by some means which does not involve any actual identification and naming. The second stage consists of a binary search of the ensemble of words belonging to this range. Treisman notes the evident difficulty, that of seeing how a word (or an object) can be allotted to a sub-group without its being identified. I will suggest that—at least in the case of an object— it is first identified only as being something having a certain degree of *familiarity*, and that this is an important early stage of perception allowing the organism to decide quickly whether it falls into a category of things which can be appropriately reacted to without more ado, or whether on the contrary more elaborate resources of identification and response-choice must be mobilized. I am not going to try and say how this could be achieved—by what cues this implicit judgement is evoked. But there *is* a connection between word-frequency, in the Thorndike-Lorge sense—and subjective judgements of frequency and familiarity. Howes (1954) showed, for words in general, that intuitive judgements of frequency of use correlated with Thorndike-Lorge entries to between 0·57 and 0·87. Wingfield (1966) found similar figures for judgements of both commonness and familiarity of the words and objects used in our naming experiments. Fraisse, Noizet and Flament (1963) obtained judgements of *familiarity* of words, by both a scaling and a paired comparison method and got a high correlation. Fraisse (1963) then measured reaction-times for reading single words aloud and found a higher correlation between reaction-time and familiarity than between the former and entries in the French word-frequency table of Gougenheim *et al.* (1956).

Let us suppose, then, an expression for the reaction-time in naming of the form:—

$$T_L = T_1 + \tau . \log_2 . n.$$

T_1 is the mean time taken in deciding in which of a relatively small number of frequency-ranges further search is to be directed. This time will be a constant for a given number of ranges, and, if the procedure is to scan the set *seriatim*, will be a linear function of that number. The second term $\tau.\log_2 n$, represents the second stage of binary search, τ being the time for each decision. We can see how far this fits the reaction-times Wingfield and I found and my count of Thorndike-Lorge, by picking out corresponding reaction-times and ensemble-sizes from the two graphs I showed, and plotting one against the other. I have had to throw away some of the data because the AA and A items do not fit on the Zipf line for the reasons I mentioned, and because I had not the hardihood to carry my object-name count below a frequency of one per million. The results of this replot are shown in Figure 4. Since both variables were picked off straight lines this also must be straight—the slight deviations are due to small errors in plotting—and we can read off the rate of processing information which is the reciprocal of the gradient, and works out at about 17 bits per sec. The Y-intercept, about 370 millisec., would presumably include not only the constant time we have supposed to be taken over the first stage of the search itself, but also that required for any previous primary processes of reception, perception and so forth. This gives an idea of the kind of argument which can be applied to this kind of data. That it is somewhat shaky is clear from the large extrapolation involved.

FIGURE 4



Naming-latency as a function of name ensemble size.

We have, of course, been thoroughly arbitrary in dividing up the part of the frequency range we used into *five* sub-ranges. Within limits, it is true, the exact number of categories of familiarity in the first part of the search would not be critical to the argument. In any case nothing in our results to date allows us to guess at the right number. But an interesting and probably important question is certainly left open. Categorization in perception is under close consideration at the moment, though much experimental work, such as that of Turner and Treisman (Personal communication) at Oxford, has to do with the *kind* rather than the *number* of categories used. We await further developments.

Identification and naming

The view I have been suggesting about the perception of an object implies that it proceeds by progressive classification or categorization: the first, or one of the first, stages being concerned with its familiarity or commonness. I will further suggest that it is quite wrong to think that there is any particular or definitive end-point to the process. It is carried just so far as is necessary for response of the kind required in the circumstances. Sometimes these requirements are of direct biological significance. Crossing a busy street I do not stop to identify a moving object as a taxi rather than a car. The coroner might be interested in this discrimination but to me it is irrelevant. Sometimes, on the other hand, identification is carried beyond the stage at which any single name-word is available, and recourse must then be had to adjectives or descriptive phrases.

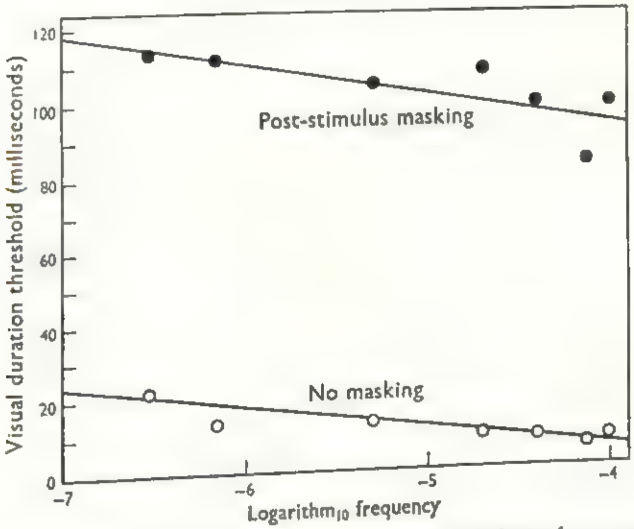
It is true, nevertheless, that the behavioural and linguistic features of our culture make the stage of naming a relatively definitive end-point in identification. We tend to invent names for things to the extent that they require differences of behaviour. Conversely, we are apt to behave towards things in accordance with the names they have—sometimes to the neglect of their properties. But it is also true that we recognize a distinction between knowing what a thing is and calling it by the right name. Animals and young children behave towards many things in much the same differentiated way that we do, without having names for them. As one gets older one is increasingly in the position of fumbling over the *name* for something, while still, in the words of the late Professor Wolters (1933), "knowing what to *do* about it."

We ought, therefore, to ask in regard to naming-latencies how much time is taken up in recognizing, or identifying, the object and how much in extracting its name from the word-store. More generally, we ought perhaps to enquire whether

the distinction which commonsense tends to make between these two phases is borne out by any available experimental data. Such data are clearly not easy to come by. In general the only evidence we are likely to get that a person has correctly identified something is provided by his telling us its name—"what it is."

Wingfield (1966), however, tried two lines of approach to this problem. The first consists in measuring the "visual duration threshold" for a person to be able to name an object. The first, simple argument on which this experiment is based would be that the identification process can only go on while the stimulus is present. If identification is not arrived at during this time it will not take place at all. In practice we have to allow a further brief but uncertain period during which not the stimulus itself, but a visual after-image or immediate memory-image remains active. However it is known from other experiments that this period can be very markedly reduced by immediate after-presentation of another, and this time meaningless, pattern. With the object pictures previously used, Wingfield obtained the results shown in Figure 5. Lines for thresholds with and without masking are shown.

FIGURE 5



Visual duration thresholds for object series as a function of name-frequency.

The visual duration threshold is very much smaller than the least of the naming-latencies, and although there is a frequency effect it is a much smaller one. So much would suggest that the "perception"—or "identification" time is but a minor part of the total latency. However I would not myself put too much faith in this argument, for it does not follow that the stimulus must necessarily be present throughout the whole of the identification time. It need only be there long enough—and at an intensity high enough—for the information required for identification to be extracted and coded. The *processing* of this information once coded might well take a further period of time. The frequency effect would simply depend on the need for more information to be extracted for rare things to be named than for common.

Wingfield's second attempt to determine the identification time is, at any rate at first sight, more promising. He did what might be described as a "Same-Different" experiment. He presented first the *name* of an object and then immediately after a *picture* of an object. The latter might be the same as, or different from, the former, and the subject's task was to say appropriately "yes" or "no" as quickly as possible.

In another experiment the test picture was preceded, not by a name but by a picture. In both cases the objects were either common, having a Thorndike-Lorge frequency of more than five in 100,000, or rare, with a frequency of less than one in 100,000. As may be seen in Table I, in which the naming latencies have been inserted for comparison, the reaction time in all these conditions was the same and statistically

TABLE I

MEAN LATENCIES IN MILLISEC. FOR THE SAME 16 OBJECTS IN NAMING AND JUDGING SAME OR DIFFERENT

	Naming	*Same-different*
Common ..	636	504
Rare ..	1169	522

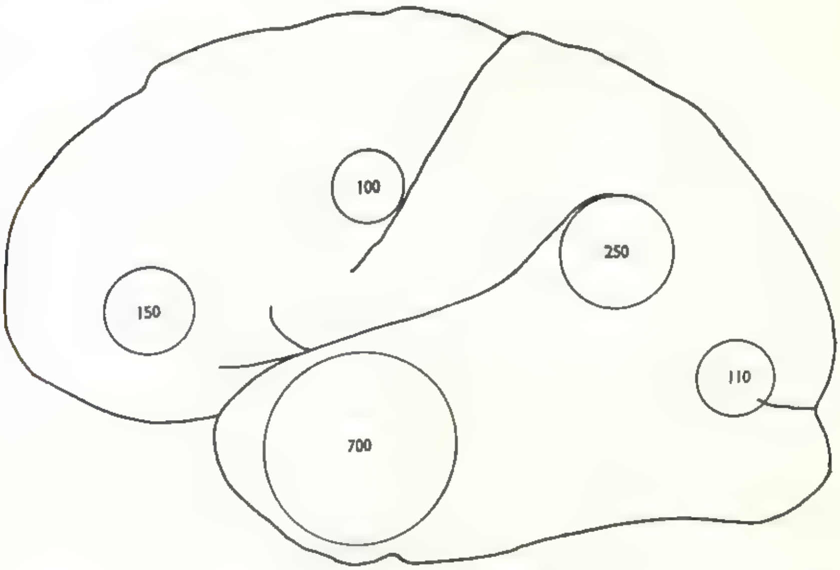
indistinguishable from 500 millisec., with the exception that for the object-object series there is a significant difference between the times for "same" and "different," the latter being the longer. Now what goes on when such judgements are made I am not going to try and say off hand, but at least it is clear that there is no frequency effect. This would be consistent with the view that a certain part of the whole process of identifying and naming takes a time which is independent of the commonness or rarity of the object and that it is only when the name has to be explicitly evoked that this factor plays a part. In the second experiment, in which both presentations are pictorial, it might, of course, be argued that the decision "same" or "different" could be made purely in terms of visual pattern without regard to representative meaning. But in the first experiment there *is* no specific first picture for comparison with the second. On the other hand, it does not seem as if the *name* is being retained for comparison with the *name* of the second object, for the evocation of the latter would presumably once again introduce a frequency effect. Wingfield's conclusion that the total latency is made up of a fixed and a variable part seems provisionally reasonable and in keeping with the outcome of our previous argument which, you will remember, led to a constant time of 370 millisec. Some preliminary work I have done recently suggests that the "same-different" technique may prove fruitful in the assessment of language disorders associated with injury or disease of the brain.

THE BRAIN'S PART IN NAMING

So far we have said nothing about the brain. The brain is a physical object having bulk and spatial parts grossly discernible. When we think of a store we generally imagine it as being organized in a spatial sense so that related items lie together in a certain region. And we also suppose that a good store has walls so that its contents are kept apart from other things. Is the word-store literally contained within, and the sole occupant of, a certain limited region of the brain? Nineteenth century writers certainly inclined towards this simple kind of view, though there were dissenters—notably Hughlings Jackson and Sigmund Freud. Today there could be less agreement on this point, if only because in the history of clinical observation an enormous number of people are known to have lost the use of this or that part of their brains, covering between them every region of the cerebral hemispheres, yet there is still surprisingly little evidence that anyone ever lost part or the whole of his store of words *as such*, though of course, for one reason or another

he may be unable to produce any *given* word appropriately either singly as in naming, or in continuous discourse. But difficulty in finding names for things is a marked feature of a variety of forms of language disturbance associated with lesions in a number of different regions of the brain. In ordinary clinical examination the patient will often in this way betray some degree of aphasia when otherwise capable of convincing continuous speech in which, indeed, he may show himself able to use names which he is incapable of producing when simply asked what an object is. He may circumlocute in his efforts to get the name to emerge and may show, for instance by referring to the use or function of the object, that he well knows what it is. It seems, in fact, as if it is the *retrieval mechanism*, rather than the *store itself*, that is at fault. Lesions of some parts of the hemispheres are more productive of such disorders than are those in others, and the findings of Hécaen and his collaborators (Marcie, P., *et al.*, 1965) for the left hemisphere shown in Figure 6 are probably representative, so far as what is apparent at the clinical level is concerned.

FIGURE 6



After Hécaen *(1964)*

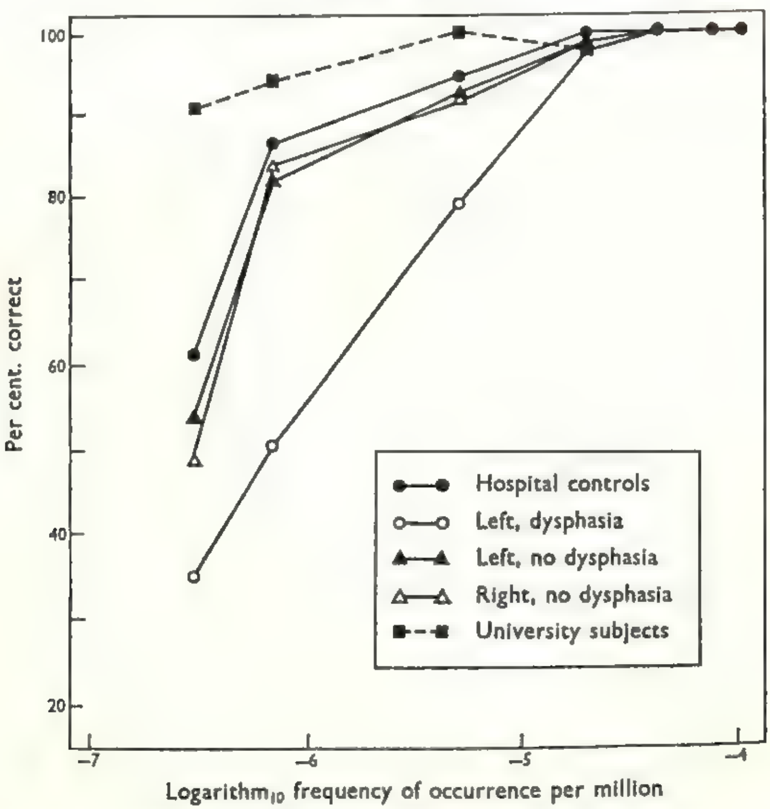
Relative incidence of naming disorders for regions of the left hemisphere—from data of Hécaen and collaborators (Marcie *et al.*, 1965).

Can we get any further than this—perhaps by using more refined techniques than the clinician has at his disposal? Lotmar as long ago as 1919 was, I think, the first clinician to publish recorded latencies, for object naming if only on one case, but it is pleasing to find that even these crude data fit a time-frequency relationship. Otherwise only Ombredane (1951), I believe, made use of latency measurements—in this case to trace oscillations in performance. But in the extensive work of Rochford and Williams (1962–5) other features were examined. They had noticed, as others must have done, that different objects present different degrees of difficulty to the aphasic. Arranging a chosen set of objects in this order, they found that substantially this same ranking applied to the age at which children could name the objects and to the difficulty—as measured by percentage success—presented at any given age. Very much the same was found with patients in a befuddled condition following electro-convulsive therapy and with normal subjects stressed by speed and

a barrage of misleading verbal stimuli. Moreover they found evidence that the *types* of error made in all these cases were similar. All this at least suggests that the disturbances found in the brain-injured are not something esoteric and peculiar to their pathological condition, but represent an exaggerated form of a partial breakdown of old-established normal function discernible in normal people.

In the course of further work Rochford established a clear frequency-effect in naming by aphasics in terms of a percentage-success measurement, not only in the straightforward task with objects but in a number of other connections, such, for instance, as the evocation of verbs in response to pictures of activities. These valuable experiments fill out and refine our ideas about naming and in many respects go beyond the limited factors with which we are concerned this afternoon. Newcombe, Wingfield and I (1965) in the course of studying latencies in a brain-injured population also obtained data in terms of percentage success. These are shown in Figure 7 and raise one or two points of interest. The first is the marked

FIGURE 7



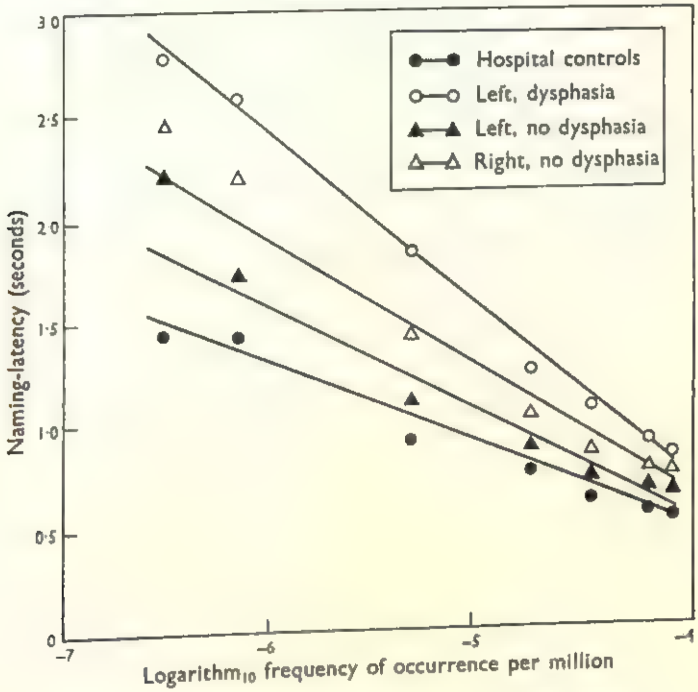
Performance at naming as a function of name-frequency. (Reprinted from Newcombe, Oldfield and Wingfield, 1965, *Nature*, **207**, 1217, by kind permission of the Editor.)

superiority of the university subjects of our first, laboratory, experiment over the supposedly normal, but perhaps less literate, hospital control group. This is a clear warning that cultural factors and personal endowment must be taken into account in this field. Secondly the non-aphasics, whether left or right hemisphere, and the normal hospital patients are not separated by the particular measurement. Lastly—here, perhaps, a hint to the clinician—the detection of mild nominal aphasia may

require the use of objects whose names occur less than about once per hundred thousand words of text.

When it comes to latencies, however, things become a little clearer. Our subjects were a group of rather over a hundred people, nearly all of whom had sustained injuries to a variety of parts of the cerebral hemispheres in the Normandy campaign of 1944 and to this extent formed a population restricted in pathology, in age, and in health before being wounded. The same set of objects as before was used, though the technique was slightly simplified to meet the different conditions of the clinical setting. A full analysis of the results is still going on—we find it better not to allow the computer to churn out too much at a time—but some interesting features have emerged. The latency-frequency relationships for these groups are shown in Figure 8. The lines are all statistically healthy and significantly different from one another.

FIGURE 8



Naming-latency as a function of name frequency.

The first point to notice is that latency-measurement *does* discriminate the four groups. As we should expect, the aphasics come out worst with the highest mean latency and largest slope—corresponding with the lowest information rate. All the brain-injured show this slowing by comparison with the normals. Is this because of some *general* slowing down of their cerebral computing machinery, or is it to be attributed specifically to a breakdown of what Goldstein called the "instrumentality" of language? A lot more work will be needed to answer this question. But the most interesting feature of these graphs is the unexpected reversal of the right and left hemisphere groups without aphasia. This might have been an artefactual consequence of dividing the left hemisphere group into those *without* aphasia and those *with*, for the latter would tend to include the larger lesions, and therefore remove them from the comparison, while the right sided group still

contains them. But this does not seem to be the explanation, for when we compare groups matched for size and severity of lesion the positions are unchanged. Now, with certain special exceptions, the whole weight of clinical evidence supports the view that in a right-handed person the left half of the brain is "dominant" and in particular contains all the language mechanisms. Surgical interference with the right, for instance, is universally held to present no hazard to speech. Indeed a certain conspiracy of silence prevails among neurologists about the lack of anything much for the right hemisphere to do. From time to time suggestions have been made about its role in speech. Hughlings Jackson, for instance, thought it might have to do with the more automatic aspects, or alternatively that it was here that words already formulated waited their turn for utterance. Much more recently Hécaen and his colleagues (Marcie *et al.*, 1965) have reported certain specifically verbal defects associated with pure right-sided lesions. The position so far as naming is concerned is that, taking a test which the clinician himself uses to establish the presence or absence of aphasia, and refining and quantifying this, we seem to discover a certain blurring in the outline of the original clinical concept. Ought we to say that the patient with a right-sided lesion *is* slightly aphasic, or ought we to account for the fact that he takes longer to name things on some other ground? The answer must partly depend on finding out the effects on naming-latencies of lesions in different *parts* of the two hemispheres respectively. Whether our own data will allow this remains to be seen.

I am afraid we cannot claim this afternoon to have taken more than a first bite—or even, perhaps, nibble—at the problem I raised. But I thought it better in the face of this audience—and particularly its principal member—to try and tackle something definite. Long experience of Sir Frederic's reaction to generalities remote from fact discourages the use of the brain to produce words without reference to things. We have tried to find out something of what goes on when words are taken out of store in response to single stimuli such as objects. What relation does this problem bear to that of word-storage and evocation in the case of continuous discourse? Little, it would seem, at first sight. The aphasic who is stumped when shown a knife and asked to name it may often generate meaningful and reasonably correct word strings, which may include this word, at a speed comparable with that of a normal person. If, on the other hand, we ask a normal person to name an object, it takes him about six-tenths of a second to do so, and an entertaining experiment is to present a succession of objects at the rate of about 100 a minute. This is well below the rate at which he produces words in ordinary talk, but does not allow enough time for each object to be named. The result is a breakdown in the flow of words, with speech blockage, naming errors and even attempted circumlocution.

Does this mean that not only the access systems, but also the storages, are different in the two cases? Perhaps this question has something in common with the old problem of whether language as a self-contained system of verbal elements inter-related by rules of use is different from language as referring to the external world of things. Certainly far more psycholinguistic research today is being devoted to the former than the latter aspect. And good progress is being made. But I do not think this robs our own more limited problem of interest and importance. It seems very unlikely that when I utter the word "pencil" in naming an object it comes from one store, while when I say "A pencil is useful because you can rub out what you write" it comes from another. But if there *is* to be a single store it must be organized not only in the way we have imagined—in terms of frequency of use or familiarity—but also in such a fashion that sentence-generating programmes can operate on it. Libraries have not only their author—but also their subject catalogues.

What this other form of organization may be we cannot say at present, but any understanding we can get of single word-extraction can scarcely fail to help us find out.

Much of the work referred to in this lecture was done by my friends and colleagues of the Medical Research Council Psycholinguistics Research Unit, especially Moyra Williams and Arthur Wingfield. Freda Newcombe, although not 'in' the Unit has been very definitely 'of' it. I cannot sufficiently express my gratitude to all of them for their leading part in the actual conduct of the experiments and for the opportunities I have had for discussion with them. Though they must not be supposed to be responsible for my theoretical interpretations, or even necessarily in agreement with them, I should never have formulated these without their stimulus and encouragement.

To the Medical Research Council I must also express our gratitude for their constant support of work at Oxford in this field over the past 10 years.

References

Cattell, J. McK. (1886). The time it takes to see and name objects. *Mind*, **11**, 63–5.

Fraisse, P. (1963). La perception des mots. In Ajuriaguerra, J. de C. *et al.*, *Problèmes de Psycholinguistique*. Paris: P.U.F.

Fraisse, P., Noizet, G., and Flament, C. (1963). Fréquence et familarité du vocabulaire. In Ajuriaguerra, J. de C. *et al.*, *Problèmes de Psycholinguistique*. Paris, P.U.F.

Goldman-Eisler, F. (1958). Speech production and the predictability of words in context. *Quart. J. exp. Psychol.*, **10**, 96–106.

Gougenheim, G., Michea, R., Rivenc, P., and Sauvageot, A. (1956). *L'Élaboration du Français Elémentaire*. Paris:

Howes, D. (1954). On the interpretation of work frequency as a variable affecting speed of recognition. *J. exp. Psychol.*, **48**, 106–12.

Howes, D., and Solomon, R. L. (1951). Visual duration threshold as a function of word probability. *J. exp. Psychol.*, **41**, 401–10.

Lotmar, F. (1919). Zur Kenntnis der erschwerten Wortfindung und ihre Bedeutung für das Denken des Aphasischen. *Schweiz. Archiv. Neurol. Psychiat.*, **5**, 206–39; **6**, 3–36.

Mandelbrot, B. (1965). Information theory and psycholinguistics. In Wolman, B. B., and Nagel, E., *Scientific Psychology*. New York: Basic Books, pp. 550–62.

Marcie, P., Hécaen, H., Dubois, J., and Angelergues, R. (1965). Les réalisations du langage chez les malades atteints de lésions de l'hemisphère droit. *Neuropsychologia*, **3**, 217–45.

Newcombe, F., Oldfield, R. C., and Wingfield, A. (1965). Object-naming by dysphasic patients. *Nature*, **207**, 1217–8.

Oldfield, R. C. (1963). Individual vocabulary and semantic currency. *Brit. J. soc. clin. Psychol.*, **2**, 122–30.

Oldfield, R. C., and Wingfield, A. (1965). Response latencies in naming objects. *Quart. J. exp. Psychol.*, **17**, 273–81.

Ombredane, A. (1951). *L'Aphasie et l'Élaboration de la Pensée explicite*. Paris: P.U.F.

Rochford, G., and Williams, M. (1962–5). Studies in the development and breakdown of the use of names. *J. Neurol. Neurosurg. Psychiat.* (I) **25**, 222–7, 1962; (II) **25**, 228–33, 1962; (III) **26**, 377–81, 1963; (IV) **28**, 407–13, 1965.

Thorndike, E. L., and Lorge, I. (1944). *The Teacher's Word Book of 30,000 Words*. New York: Columbia University Press.

Thumb, A., and Marbe, K. (1901). *Experimentelle Untersuchungen über die psychologischen Grundlagen der sprachlichen Anologiebildung*. Leipzig: Engelmann.

Treisman, Anne M. (1965). Effect of verbal context on latency of work selection. *Nature*, **206**, 218–9.

Wingfield, A. (1966). *The Identification and Naming of Objects*. D.Phil. Thesis, Oxford.

Wolters, A. W. P. (1933). On conceptual thinking. *Brit. J. Psychol.*, **24**, 133–43.

Zipf, G. K. (1935). *The Psychobiology of Language*. Boston: Houghton Mifflin.

SHORTER ARTICLES AND NOTES

STIMULUS VARIATION AND SEQUENTIAL JUDGEMENTS OF DURATION

BY

G. VON STURMER

From the Department of Psychology, Monash University

When a series of reproductions of an interval is made in the absence of a standard the judgements progressively lengthen. The similarity between stimulus conditions in this type of time estimation experiment and the conditions which produce a decrement in human vigilance is discussed. It is argued that failure to detect cues for the passage of time reduces the amount of time perceived to elapse. Reproduced judgements must consequently be increased in length to match remembered standards. The hypothesis is then made that the kind of variation in background stimulation which facilitates vigilance should increase the frequency of detection of cues for duration and reduce reproduced judgements. This hypothesis is tested with 80 subjects and a reversal of the serial reproduction effect is found on trials with changed background conditions.

Introduction

When subjects are required to make a series of reproductions of an interval of time, without presentation of the standard interval between judgements, the estimates progressively lengthen (Brown and Hitchcock, 1965; Falk and Bindra, 1954). Using the method of comparison, Woodrow (1935) observed a similar effect. Treisman (1963) has suggested that the effect may result from a decline in "specific temporal arousal" which, in his model for an internal clock, affects a pacemaker providing impulses for a counter.

Since it is common in experiments of this sort to keep extraneous stimulation as monotonous as possible, effects similar to those found in studies of human vigilance may be expected (Buckner and McGrath, 1963). As the experiment progresses the subject may have increasing difficulty in detecting cues from which to infer the passage of time. Consequently he must allow the timer to run on further and further to make judgements of apparently equal durations.

In a review of the literature on vigilance, Broadbent (1964) suggests that variations in background stimulation which produce a moderate increase in arousal in drowsy subjects may improve performance. McGrath (1963), for example, found that, when vigilance signals were presented at a slow rate, varied background stimulation in one modality improved vigilance in another. In several experiments on time estimation Frankenhaeuser (1959) found that when the background was varied verbal estimations were larger and operative judgements were smaller, suggesting that more time was perceived to elapse under conditions favouring low detection thresholds.

The aim of the experiment reported here was to examine the effect of a variation in background stimulation (*alerting stimulus*) introduced during a series of reproductions of an interval of time. It was expected that, as in the studies by Brown and Hitchcock, and Falk and Bindra, the judgements would progressively lengthen (Hypothesis 1).

It was predicted that if this serial effect were due to an increasing difficulty in detecting time-relevant cues, the interpolation of a change in background stimulation would, at least temporarily, reverse the effect and lead to shorter judgements (Hypothesis 2).

It was also argued that pre-exposure to an alerting stimulus would decrease its power to reduce the serial effect (Hypothesis 3). This hypothesis was based on Berlyne's (1960) views on the role of novelty in the elicitation of an arousal reaction.

Method

Procedure

Each subject made 20 reproductions of a 5-sec. interval which was demonstrated five times before the series of judgements began. Forty subjects received an alerting stimulus

during their 5th and 6th estimates (Early) and 40 while making their 17th and 18th estimates (Late). Twenty subjects in each of these groups received an auditory alerting signal (Bell) and 20 in each group received a visual signal (Light). For 10 subjects in each of the above treatment combinations an attempt was made to reduce the novelty of the light or the bell by pre-exposure to them before testing began. These subjects were members of the low novelty group (Low). The remaining 10 subjects in each group received no pre-exposure to the alerting stimuli (High). Different treatment combinations were thus given to eight groups with 10 subjects in each group.

Subjects sat at a table facing an 8 in. by 15 in. translucent screen in a semi-dark room. Before testing, subjects in the Bell-Low group were presented with four 2-sec. rings separated by 1-sec. intervals. Subjects in the Light-Low group were exposed to 10 0·5 sec. flashes with 0·5 sec. between flashes. These events were controlled manually and timed with a stopwatch. The bell was a mechanical doorbell. The light stimulus was provided by a slide projector mounted on the other side of the screen facing the subject. The projector was operated through an electronic timer set to give a 0·5 sec. impulse whenever the experimenter pressed a remote control button. Since subjects in a pilot study had complained that the light was painfully intense two superimposed slides were placed in the projector to give an irregularly shaped dark patch in the centre of the screen.

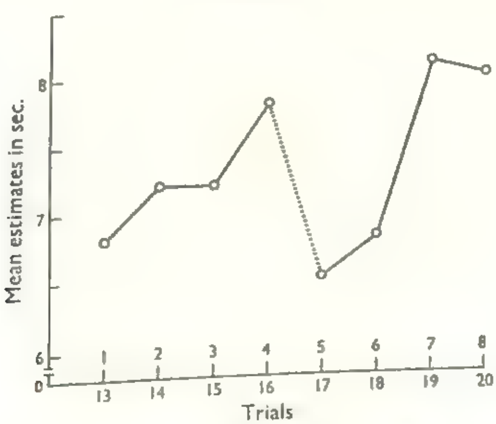
Following either pre-exposure or a 3-min. wait subjects were shown how to use a switch to start and stop a recording timer the dials of which were visible only to the experimenter. The subject was then told not to count or tap and to attend carefully. The 5-sec. standard was presented, its boundaries being marked by the click of relays in a timer similar to the one controlling the projector. After five demonstrations of the standard the subject made 20 reproductions of it. The interval between judgements was determined by the time it took the experimenter to note the interval recorded to 0·01 sec. on the response timer, reset the timer and say: "and again." The alerting stimuli were presented for the durations of the judgements on either the early or the late trials as described above.

Subjects

The subjects were 80 volunteers from 1st and 2nd year classes and the junior staff in the Department of Psychology at the University of Melbourne. None of the subjects knew the aims of the study.

Results. Results from all groups were very similar. In Figure 1, in which data from the first eight trials of the Early group are combined with data from trials 13 to 20 of the Late group, the serial effect and the downward deflection of alerted trials are clearly illustrated.

FIGURE 1



Combined data from Early and Late groups showing the serial effect and the effect of alerting stimuli. The dotted line shows the initial fall in estimates from non-alerted to alerted conditions. The Early alerted trials were trials 5 and 6. The Late alerted trials were trials 17 and 18.

The raw scores were examined in the light of Boneau's (1960) conclusions about the robustness of the t test. There appeared to be no reason to doubt the direct applicability of the F ratio and t test to the untransformed data. In the following analyses two-tailed tests were used.

Before the significance of the drop in judgement magnitudes on the alerting trials was examined, effects due to type of alerting stimulus, level of novelty (pre-exposure) and early or late presentation were sought. For each subject the estimate on the trial upon which the alerting stimulus was first given was subtracted from the score on the previous trial. These data were then subjected to an analysis of variance for a $2 \times 2 \times 2$ factorial design (Winer, 1962). None of the effects reached the 0·05 level of significance. Hypothesis 3 was thus not confirmed.

Since the Bell-Light, High-Low and Early-Late manipulations led to no significant differences in the degree or direction of change in estimates from non-alerted to alerted conditions, it became possible to pool data from these sources for a test of the alerting effect itself on all 80 subjects. This was achieved by comparing the mean estimate (6·5 sec.) on the first alerted trial with the mean estimate (7·8 sec.) on the immediately preceding trial using a t test for correlated groups. The difference was significant ($t = 5{\cdot}5$, $d.f. = 79$, $p < 0{\cdot}001$). Hypothesis 2 was thus confirmed.

It was then necessary to show that the decrease in estimates on alerted trials was embedded in a series of trials exhibiting an opposing trend—the serial effect. Eighteen trials were available for testing the serial effect and an *a priori* decision was taken to analyse eight of these, four trials bracketing the Early alerted trials (trials 1, 4, 7 and 8) and four trials bracketing the Late alerted trials (trials 13, 16, 19 and 20). In this way the selected trials were spread economically and symetrically over the available range.

Data from the Bell-Light and High-Low categories were pooled. A 2×4 factorial analysis of variance was used with repeated measures on one factor (Winer, 1962). Early-Late was factor A, treated at two levels, a_1 and a_2. Trials gave the repeated measures factor, factor B, which was treated at four levels. With this design it was expected that the serial effect over trials on either side of the alerted trials would appear as a significant effect on factor B. The serial effect between Early and Late samples would show on factor A. An $A \times B$ interaction could be taken to mean either that the serial effect was asymptotic or that the alerting stimuli on trials 17 and 18 also depressed scores on trials 19 and 20. Table I shows that the results of this analysis supported Hypothesis 1.

TABLE I

RESULTS OF ANALYSIS OF VARIANCE TO TEST FOR SERIAL EFFECT

Source	*S.S.*	*d.f.*	*F*	*p*
Between *Ss*	8130·88	79	—	—
A (Early/Late)	572·72	1	5·91	<0·05
Subj. w/in groups	7558·16	78	—	—
Within *Ss*	708·67	240	—	—
B (Trials)	86·69	3	11·28	<0·001
AB	21·69	3	2·82	<0·05
B × Subj. w/in groups	600·29	234	—	—

A further test of Hypothesis 3 was made by comparing data from the first and second presentations of the alerting stimuli (trials 5 versus 6, 17 versus 18). Data from the Bell and Light categories were pooled and the High-Low and Early-Late distinctions were kept in a factorial $2 \times 2 \times 2$ design for repeated measures on one factor (presentations). No effects reached significance at the 0·05 level.

DISCUSSION

The data from this experiment demonstrate that there was a significant tendency for the judgements to get longer as the trials progressed. Judgements on alerted trials, however, were significantly shorter than those on the immediately preceding non-alerted

trials. No significant effects attributable to pre-exposure or repeated stimulation were found. The results support the general notion that extended series of perceptual judgements in monotonous conditions are influenced by some of the determinants of the vigilance decrement. This should be considered in the design of psychophysical studies where similar conditions are used and averages of large numbers of judgements are required.

More particularly, support is given to the hypothesis that the threshold for the detection of cues for the lapse of time is raised by monotonous stimulation and lowered by varied stimulation. No conclusive statement about the role of stimulus novelty can be made, however, because this variable was not explored over a sufficiently wide range of levels.

This report is based on research submitted as a partial requirement for the degree of Master of Arts in the University of Melbourne.

References

BERLYNE, D. E. (1960). *Conflict, Arousal and Curiosity.* New York: McGraw-Hill.

BONEAU, C. A. (1960). The effects of violations of assumptions underlying the *t* test. *Psychol. Bull.*, **57**, 49–64.

BROADBENT, D. E. (1964). Vigilance. *Brit. Med. Bull.*, **20**, 17–20.

BROWN, D. R., and HITCHCOCK, L. (1965). Time estimation: dependence and independence of modality-specific effects. *Percept. Mot. Skills*, **21**, 727–34.

BUCKNER, D. N., and MCGRATH, J. J. (Ed.) (1963). *Vigilance: a Symposium.* New York: McGraw-Hill.

FALK, J. L., and BINDRA, D. (1954). Judgement of time as a function of serial position and stress. *J. exp. Psychol.*, **47**, 279–82.

FRANKENHAEUSER, M. (1959). *Estimation of Time: an Experimental study.* Stockholm: Almquist & Wiksell.

MCGRATH, J. J. (1963). Irrelevant stimulation and vigilance performance. In BUCKNER, D. N., and MCGRATH, J. J. (Ed.), *Vigilance: a Symposium.* New York: McGraw-Hill, pp. 3–21.

TREISMAN, M. (1963). Temporal discrimination and the indifference interval: implications for a model of the "internal clock." *Psychol. Monogr.*, 77, Whole No. 576.

WINER, B. J. (1962). *Statistical Principles in Experimental Design.* New York: McGraw-Hill.

WOODROW, H. (1935). The effect of practice upon time-order errors in the comparison of temporal intervals. *Psychol. Rev.*, **42**, 127–52.

Manuscript received 21st April, 1966.

ERRORS OF COMMISSION AS A FUNCTION OF AGE AND TEMPERAMENT IN A TYPE OF VIGILANCE TASK

BY

G. S. TUNE

From the Medical Research Council, Unit for Research on Occupational Aspects of Ageing, Department of Psychology, University of Liverpool

Forty subjects monitored a 40 min. series of 10-sec. intervals containing digits (spoken at the rate of 1 per sec.), each followed by 10-sec. silence. The task was to report whether or not three consecutive and different odd digits occurred. Responses were forced. The results showed that there was no correlation between either age or temperament and the number of correct detections made. Older subjects, however, made more errors of commission, and were less able to distinguish wanted from unwanted events. The younger and introverted subjects appeared to be more cautious. The data is discussed in terms of the arousal theory of vigilance performance.

Introduction

Griew and Lynn (1960) have argued that older and extraverted subjects should be worse at vigilance tasks than younger and introverted subjects because they generate reactive inhibition faster and dissipate it more slowly. While it is well established that introverts do make more correct detections than extraverts (Bakan, 1959; Bakan, Belton and Toth, 1963), the evidence for an age difference in correct detections is not unanimous. Davies and Griew (1965) in a survey of the topic indicated that old people achieved similar detection scores to younger people. Surwillo and Quilter (1964), however, did show an age difference in performance in the Mackworth Clock test, but this emerged only after 45 min. on the task.

It has also been reported (Tune, 1966) that while older and younger subjects made comparable correct detection scores in a forced-choice response vigilance task, the former made many more errors of commission (reported many unwanted events as if they were in fact "wanted"). In an extreme case the effect of this response bias would be that subjects achieved a maximum correct detection score but also a maximum number of commissive errors. Obviously the more errors of this type a subject is prepared to make then the greater is the insurance against an error of omission. Such a response strategy would have the effect of masking, either partially or wholly, any individual differences which might be expected in the correct detection score. The question remains, however, as to whether or not there are temperament differences in the number of errors of commission subjects make. If the vigilance performance of older and extraverted subjects is comparable, as Griew and Lynn (1960) predicted, it is likely that these groups would make relatively large numbers of commissive errors. It is further likely that both age and temperament differences would not be apparent in the correct detection scores.

The present experiment was designed to examine these points in a vigilance task where forced-choice responses were required, and where errors of commission were not discouraged, where the performance of young, old, introverted and extraverted subjects could be compared.

Procedure

Subjects listened to a tape recording of groups of 10 digits (the numbers 1 to 9 inclusive were used in a quasi-random order) each group being followed by 10 sec. silence. The task was to identify three consecutive, different and odd digits (e.g. 517, 93 5, etc.) occurring somewhere in the group (the exact location of the three "wanted" digits was randomized such that neither the first nor the last number in any group of 10 was the beginning or end of an odd-odd-odd triad). No more than one "wanted" triad occurred in any one block of 10 digits. Subjects were instructed to listen to all 10 numbers and then to place a tick in one of two columns on a prepared answer sheet. The two columns were headed "No" (a wanted combination of numbers did not occur) and "Yes" (a wanted combination did occur). In the latter case subjects were also encouraged to write the triad in a third column: it was emphasized, however, that this was unimportant

(transcription errors were not counted against a subject). Scores were assessed entirely in terms of the number of ticks placed in the "Yes" column. The subjects listened to 120 groups of 10 digits over a period of 40 min.: 27 of these contained a "wanted" triad. A brief practice session consisting of two groups of 10 digits (one with and the other without a "wanted" triad) was given in order to ensure that subjects understood the procedure. The experimenter remained with the subject for the first $\frac{1}{2}$ min. of the task and then left.

Subjects. Forty subjects were used. These were classified as old (those over 50 years) or young (those of 49 years and under). They were further classified into introverts or extraverts by means of their score on Part 2 of the Heron (1956) inventory. A division was taken at the median score to ensure equal groups. There were thus four groups of 10 subjects (see Table I). In addition each group contained five men and five women.

Results

Two response measures were taken from the raw data produced by each subject: these were (*a*) the number of "Yes" responses which were correct detections and (*b*) the number of "Yes" responses which were errors of commission. The distribution of "Yes" responses by type, age and temperament is shown in Table I. It is apparent that in the four groups the probability of making a correct detection was about the same. The error of commission scores, however, show a different picture. Spearman rank correlations were calculated therefore between the two response measures and the two independent variables and are given in Table II. These indicate that the older subjects made more

TABLE I

The Distribution of "Yes" Responses by Age, Temperament and Type

	N	*Mean age*	*Mean Heron 2 score*	*Correct detections*	*Errors of commission*
Old Introverts ..	10	65·9	7·8	248	96
Old Extraverts ..	10	65·3	3·9	234	325
Young Introverts ..	10	35·5	9·3	246	16
Young Extraverts ..	10	38·1	4·4	252	32
Totals	40	—	—	980	469

TABLE II

Spearman Rank Correlations Between the Response Measures and the Independent Variables

	ρ	$p <$ (*two-tailed*)
Age × Correct Detections ..	−0·046	NS
Age × Errors of Commission ..	+0·521	0·001
Age × d′	−0·498	0·01
Age × β	−0·417	0·01
Heron (2) × Correct Detections ..	+0·021	NS
Heron (2) × Errors of Commission	−0·352	0·05
Heron (2) × d′	+0·281	0·10
Heron (2) × β	+0·395	0·01

errors of commission whereas there was no relationship between age and correct detection score, although there was a slight tendency for the older people to make fewer (as shown by the negative sign of the correlation). Temperament was not significantly associated with correct detections although there was a tendency for high Heron 2 scorers

(introverts) to make more. A low Heron 2 score on the other hand was significantly associated with a high error of commission score.

The probability of making a correct detection and the probability of making an error of commission were calculated for each subject and from these scores d′ and β were ascertained by reference to tables (Freeman, 1964). If the former is taken as a measure of signal discriminability and the latter as a measure of caution (Broadbent and Gregory, 1963), it is apparent from the correlations between these scores and the independent variables (Table II) that the older people and the extraverts were less able to discriminate signals from non-signals whereas the younger people and introverts were more cautious.

Discussion and Conclusions

It seems clear that although, in this experiment, there were no individual differences in correct detections, the older and extraverted subjects did exhibit similar performance. Increasing age and extraversion were both associated with large error of commission scores, an inability to distinguish wanted from unwanted events and a tendency to be incautious.

These results are apparently in agreement with some findings, that there is no age difference in detection scores (Davies and Griew, 1965), and apparently in conflict with others, that there are temperament differences in correct detection scores (Bakan, 1959). If errors of commission are considered, however, the picture is changed. It was emphasized in the introduction that such errors would tend to obscure individual differences in detection scores. This, it seems, is what happened here although the older and extraverted subjects tended to detect fewer wanted events than the younger and introverted.

The problem of why some groups made more errors of commission than others is not easy to answer. Although these data lend support to Griew and Lynn's reactive inhibition hypothesis, other accounts of vigilance performance must be considered.

Broadbent (1963) has argued that introverts are chronically aroused, hence extraverts may be regarded as underaroused. While there is little physiological evidence to support this view, it has been found useful in explaining temperament differences in vigilance performance (Broadbent, 1963). It is tempting to assume further, that older people are likewise underaroused (Tune, 1966). From this point of view it would be reasonable to infer that underaroused subjects would attempt to raise their arousal level to a suitable level by increasing their response information (alternating between the two types of response, "Yes" and "No," more readily). This would result in the underaroused groups making more commissive errors, as was found here. On the other hand, Welford (1965) has argued cogently that older subjects are in fact overaroused. What the effect of such a state would be on such subjects' error of commission score is not entirely clear. It is possible that overarousal might imply overactivity which could be exhibited in a tendency to alternate between response categories. If this was accepted, it would mean that over and underarousal were behaviourally indistinguishable since both would produce high numbers of commissive errors. For this reason it is perhaps more economical to regard older subjects as underaroused.

A second approach is to avoid speculation about arousal states, since the evidence regarding older people in this respect is not unequivocal, and to posit that older and extraverted subjects are inherently less cautious. This view implies that such groups could insure against low detection scores by reporting many events as wanted even when they were not. The fact that β, in this experiment, was negatively correlated with age, and positively with a high Heron 2 score (introversion) lends some support to the argument, although it conflicts with Craik (1965) who found that older subjects were more cautious than younger. These two approaches are not necessarily unrelated, for it is likely that arousal level and cautiousness are not independent of each other (Welford, 1962).

From the results of this experiment and what has been said some speculations may be made. The arousal approach seems to imply that errors of commission can be regarded as an index of need for information. This ties in well with the "sensoristasis" hypothesis (Schultz, 1965) and implies that subjects will seek information in order to perform adequately. Errors of commission should, therefore, increase with time spent on a progressively de-arousing task. Commissive errors should also distinguish between differential levels of arousal. The present data show that age correlated with errors of commission at a higher level of significance than did temperament. This is only slight

evidence that age is more de-arousing than extraversion and more refined evaluations of this point need to be made.

I should like to thank Mr. R. Aldridge-Morris for carrying out some of the experimental work described here.

References

BAKAN, P. (1959). Extraversion—Introversion and improvement in an auditory vigilance task. *Brit. J. Psychol.*, **50**, 325–32.

BAKAN, P., BELTON, J. A., and TOTH, J. C. (1963). Extraversion—Introversion and decrement in an auditory vigilance task. In BUCKNER, D. N., and MCGRATH, J. J. (Ed.), *Vigilance: A Symposium.* New York: McGraw Hill. Pp. 22–8.

BROADBENT, D. E. (1963). In BUCKNER, D. N., and MCGRATH, J. J. (Ed.), *Vigilance: A Symposium.* New York: McGraw-Hill. Pp. 192.

BROADBENT, D. E., and GREGORY, M. (1963). Vigilance considered as a statistical decision. *Brit. J. Psychol.*, **54**, 309–23.

CRAIK, F. I. M. (1965). *Age Differences in Confidence and Decision Processes.* Unpublished Ph.D. Thesis, University of Liverpool.

DAVIES, D. R., and GRIEW, S. (1965). Age and vigilance. In WELFORD, A. T., and BIRREN, J. E. (Ed.), *Behavior, Aging and The Nervous System.* Springfield: Thomas. Pp. 54–9.

FREEMAN, P. R. (1964). Table of d′ and β. *Med. Res. Coun., A.P.R.U.*

GRIEW, S., and LYNN, R. (1960). Construct "Reactive Inhibition" in the interpretation of age changes in performance. *Nature*, **186**, 182.

HERON, A. (1956). A two part personality inventory for use as a research criterion. *Brit. J. Psychol.*, **47**, 243–51.

SCHULTZ, D. P. (1965). *Sensory Restriction Effects on Behavior.* New York: Academic Press.

SURWILLO, W. W., and QUILTER, R. E. (1964). Vigilance, age and response time. *Amer. J. Psychol.*, **77**, 614–20.

TUNE, G. S. (1966). Age differences in errors of commission. (Unpublished paper.)

WELFORD, A. T. (1962). Arousal, channel-capacity and decision. *Nature*, **194**, 365–6.

WELFORD, A. T. (1965). Performance, biological mechanisms and age: a theoretical sketch. In WELFORD, A. T., and BIRREN, J. E. (Ed.), *Aging, Behavior and the Nervous System.* Springfield: Thomas. Pp. 3–20.

Manuscript received 10th June, 1966.

SHORT-TERM MEMORY FOR WORD SEQUENCES AS A FUNCTION OF ACOUSTIC, SEMANTIC AND FORMAL SIMILARITY

BY

A. D. BADDELEY

From the Medical Research Council, Applied Psychology Research Unit, Cambridge, England

Experiment I studied short-term memory (STM) for auditorily presented five word sequences as a function of acoustic and semantic similarity. There was a large adverse effect of acoustic similarity on STM (72·5 per cent.) which was significantly greater ($p < 0{\cdot}001$) than the small (6·3 per cent.) but reliable effect ($p < 0{\cdot}05$) of semantic similarity.

Experiment II compared STM for sequences of words which had a similar letter structure (formal similarity) but were pronounced differently, with acoustically similar but formally dissimilar words and with control sequences. There was a significant effect of acoustic but not of formal similarity.

Experiment III replicated the acoustic similarity effect found in Experiment I using visual instead of auditory presentation. Again a large and significant effect of acoustic similarity was shown.

Introduction

In a series of short-term memory (STM) experiments Conrad (1963, 1964) has shown that sequences of items which are hard to discriminate in noise are also hard to remember, even though presented visually. Analogous effects of intra-list similarity have also been shown in long-term memory (LTM) where several types of similarity have proved to be relevant including similarity of letter structure (Horowitz, 1961) and of meaning (Underwood and Goad, 1951; Baddeley and Dale, 1966). However, Baddeley and Dale (1966) using paired-associate learning failed to show an equivalent effect of semantic similarity in STM and suggested that STM may differ from LTM in relying more on acoustic cues and much less on the meaning of the material to be retained. The present study uses the method of serial recall to explore further the role of similarity in STM.

Experiment I compares the influence of acoustic similarity on ordered STM for word sequences with that of semantic similarity.

Experiment I

Method

Design

A separate group of subjects did each of two conditions, A and B. Both groups attempted to recall 24 sequences of five words. In condition A these comprised 12 drawn from a set of eight acoustically similar words (mad, man, mat, cap, cad, can, cat, cap) and 12 from a control set of acoustically different words of equal Thorndike-Lorge frequency (Thorndike and Lorge, 1944) (cow, day, bar, few, hot, pen, sup, pit). In Condition B, 12 sequences were drawn from a set of eight adjectives with similar meanings (big, long, broad, great, high, tall, large, wide) and 12 from a set of eight semantically different words of equal Thorndike-Lorge frequency (old, deep, foul, late, safe, hot, strong, thin). All sequences were drawn at random with the constraint that no word appeared more than once in the same sequence. Similar and different sequences were presented in the same random order in both conditions.

Procedure

Subjects were tested in groups of about 20. Word sequences were presented by tape recorder at a rate of one word per sec. and subjects were allowed 20 sec. to write their ordered responses. To maximize response availability the relevant sets of words were written on cards and both sets were visible throughout the test session. Subjects were instructed that no sequence would contain words from more than one set but were not told before each sequence which set would be involved. A listening test was given both

before and after the memory test to ensure that subjects were hearing words correctly. The 16 relevant words were presented in random order and subjects were allowed 5 sec. per word to copy them. Two Condition A subjects did not score perfectly and were discarded, leaving 20 subjects in Condition A and 21 in Condition B. Housewives from the A.P.R.U. subject panel served as subjects. These were paid for participation and were assigned haphazardly to one of the two conditions.

Results

Performance was scored in terms of percentage correct sequences, and since scores were not normally distributed they were analysed using non-parametric tests.

Acoustic similarity. A mean of 9·6 per cent. of the acoustically similar sequences were correctly reproduced (range 0–33·3), compared with 82·1 per cent. of the control sequences (range 58·3–100). Since there is no overlap between the two distributions, the difference is clearly highly statistically significant, $p < 0{\cdot}001$.

Semantic similarity. The mean score for semantically similar sequences was 64·7 per cent. correct (range 16·7–100), and that for control sequences was 71·0 per cent. (range 16·7–100). Although the mean difference is only 6·3 per cent. a Wilcoxon test indicates that it is statistically significant, $p < 0{\cdot}05$.

Comparing acoustic and semantic similarity. The mean difference between acoustically similar and control sequences was 72·5 per cent. (range 50·0–91·7). The equivalent effect of semantic similarity was 6·3 per cent. (range 0–41·7). Since there is no overlap between the two distributions, the greater effect of acoustic similarity is clearly statistically significant, $p < 0{\cdot}001$.

These results suggest that STM for word sequences shows a massive effect of intra-sequence acoustic similarity compared with only a slight effect of semantic similarity. In fact, however, Experiment I confounds acoustic and formal similarity, since the words which sound alike do so because they have letters in common. Since formal similarity has been claimed to have a marked effect on verbal learning (Horowitz, 1961) it is clearly desirable to separate its effect on STM from that of acoustic similarity. Experiment II attempts to do this.

Experiment II

Method

Design

All subjects attempted to recall 24 five-word sequences, comprising eight sequences drawn at random by sampling without replacement from each of three sets of words. Set A comprised five words which were acoustically similar but relatively dissimilar in letter-structure (bought, sort, taut, caught, wart), Set B comprised five words with similar letter structure but relatively dissimilar pronunciation (rough, cough, through, dough, bough) and Set C comprised five words of approximately equal Thorndike-Lorge frequency to sets A and B, presenting a roughly equivalent degree of spelling difficulty due to unusual letter structure or occurrence of homophones (e.g. caught–court, dough–doe), but which both sound and look relatively dissimilar (plea, friend, sleigh, row, board).

Procedure

The 24 sequences were presented in random order to a group of 17 housewives. As in Experiment I presentation was auditory, the rate was one word per sec. and subjects were allowed 20 sec. to write their responses. To maximize response availability the three sets of words were written on cards which were visible throughout the test session. To prevent the use of position on the card as a cue, four cards were used, each with a different order and were interchanged frequently. Again listening tests were given before and after the main experiment. In each test, the 15 words were read out in random order and the subject was allowed 5 sec. to write down what she heard. All 17 subjects scored perfectly on both tests.

Results

Mean recall scores were as follows, acoustically similar words 36·5 per cent. correct sequences (range 0–100), formally similar 55·8 per cent. (range 25·0–100), control words 63·5 per cent. (range 12·5–100). Comparison using the Wilcoxon test showed that performance on acoustically similar sequences was significantly poorer than performance on either control sequences, $p < 0{\cdot}001$, or formally similar sequences, $p < 0{\cdot}01$. There

was no significant difference between performance on formally similar and control sequences, $p > 0{\cdot}05$.

The general level of performance differs from Experiment I in being higher for acoustically similar sequences but lower for control sequences. The acoustically similar sequences are probably easier for two reasons. First, they are selected from a set of five instead of eight items, which seems likely to reduce both degree of inter-item confusion and information load. Secondly the items differ only in terms of the initial sound, so that if the subject remembers only the initial letter, she can easily reconstruct the sequence. This latter point also holds for the other two word sets since neither has more than one word starting with the same letter, although reconstructing the control words might take slightly longer since the latter part of these words have neither a common sound nor a similar letter structure. The fact that the words used in this study were less frequent than those used in Experiment I and presented more difficulties due to spelling and competition from homophones probably accounts for the rather lower performance on control sequences.

Although the set of acoustically similar words are more formally similar than would be ideal, and the formally similar words are probably not as acoustically distinctive as the control list, it nevertheless seems fairly clear that acoustic rather than formal similarity was the principal source of difficulty.

However, both Experiment I and Experiment II have used auditory presentation so that both results are open to the objection that some of the acoustic effect may be due to perceptual error. The listening tests used suggest that it is unlikely that very much of the acoustic similarity effect is due to mishearing, but the possibility exists that some of the effect may be due to the interaction of perceptual and memory loads. The following experiment therefore attempts to replicate the acoustic similarity effect using visual presentation.

Experiment III

Method

Material

Twenty-four five-word sequences were prepared comprising 12 sequences drawn at random from a set of 10 acoustically similar words (mad, man, map, mat, max, can, cad, cap, cat, cab), and 12 from a control set (pen, rig, day, bar, cow, sup, pit, hot, few, bun). Each word was typed on a $4\frac{1}{2} \times 3\frac{1}{2}$ in. card.

Procedure

Subjects were tested individually; the experimenter presented the material manually at a rate of one word per sec. after which the subject attempted to write down the sequence in the appropriate order. The two sets of words were visible throughout the experiment to maximize response availability, and again several different arrangements of each set were used. Ten young enlisted men served as subjects.

Results

The mean recall score for acoustically similar sequences was 1·7 per cent. (range 0–8·3) and for control sequences was 58·3 per cent. (range 8·3–91·7). The clear difference between the two types of sequence was shown by all 10 subjects and is thus highly statistically significant, $p < 0{\cdot}001$.

The overall level of performance was lower than that shown in Experiment I. The question of whether this is due to the different method of presentation, the selection of sequences from 10-word instead of eight-word sets or to the different type of subject used is, however, beyond the scope of the present study.

Discussion

All three experiments agree in showing a large and consistent adverse effect of acoustic similarity on ordered STM for words, and Experiments I and II show that neither semantic nor formal similarity has an effect of comparable magnitude. The relative unimportance of semantic similarity shown in Experiment I together with the failure of Baddeley and Dale (1966) to find an effect of semantic similarity among stimuli on STM for paired associates suggests that subjects show remarkable consistency and uniformity in using an almost exclusively acoustic coding system for the short-term remembering of disconnected words. There is abundant evidence that this is not true of LTM (Underwood, 1951; Underwood and Goad, 1951; Baddeley, 1966; Baddeley and Dale, 1966).

References

Baddeley, A. D. (1966). The influence of acoustic and semantic similarity on long-term memory for word sequences. *Quart. J. exp. Psychol.*, **18**, 302–9.

Baddeley, A. D., and Dale, H. C. A. (1966). The effect of semantic similarity on retro-active interference in long- and short-term memory. *J. verb. Learn, verb. Behav.* (in press).

Conrad, R. (1963). Acoustic confusions and memory span for words. *Nature*, **197**, 1029–30.

Conrad, R. (1964). Acoustic confusion and immediate memory. *Brit. J. Psychol.*, **55**, 75–84.

Horowitz, L. M. (1961). Free recall and ordering of trigrams. *J. exp. Psychol.*, **62**, 51–7.

Thorndike, E. L., and Lorge, I. (1944). *The Teacher's Word Book of 30,000 Words*. New York: Teachers' College, Columbia University.

Underwood, B. J. (1951). Studies of distributed practice: II. Learning and retention of paired-adjective lists with two levels of intra-list similarity. *J. exp. Psychol.*, **42**, 153–61.

Underwood, B. J., and Goad, D. (1951). Studies of distributed practice: I. The influence of intra-list similarity in serial learning. *J. exp. Psychol.*, **42**, 125–34.

Manuscript received 22nd February, 1966.

OVER-CONSTANCY AND VISUAL ACUITY

BY

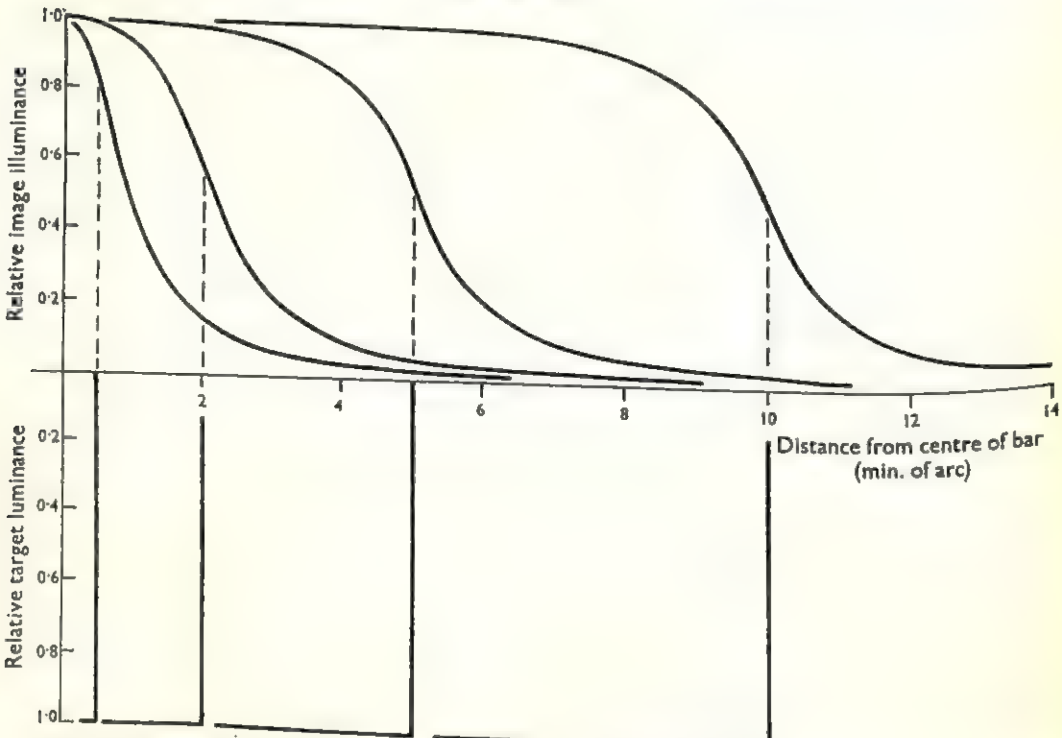
R. W. GUBISCH

From the Physiological Laboratory, Cambridge

A recent study by Joynson, Newson and May (1965) has demonstrated that errors of judgement in estimating the sizes of remote objects are closely correlated with the visual angles subtended by those objects; they observed that the errors, overestimations which increased as the angles decreased, disappear above about two degrees. It was suggested that the diameter of the fovea was involved in these results. However, the majority of their measurements lie in that range of angles for which the human eye is a poor optical instrument, adding significant amounts of blur to an already small image (Campbell and Gubisch, 1966). It is therefore important to consider here the possible influences of visual acuity upon size-constancy measurements.

If the human eye were optically perfect, its retinal images would be exact but reduced reproductions of observed objects; sharp edges in the external world would remain equally sharp as images. Because the eye is imperfect for even small pupil diameters (Arnulf and Dupuy, 1960; Campbell and Green, 1965), the abrupt edge between black and white becomes blurred into a smooth transition which makes location of the original boundary impossible. The curves of Figure 1 demonstrate this effect. They represent retinal

FIGURE 1



Blurred retinal images corresponding to white bars of widths 1, 4, 10 and 20 minutes of arc on a black background. Observer is refracted perfectly with a 4 mm. diameter pupil.

light distributions arising from rectangular white bars on a black background, similar to the planks used by Joynson *et al.* The curves were calculated from the results of Campbell and Gubisch (1966).

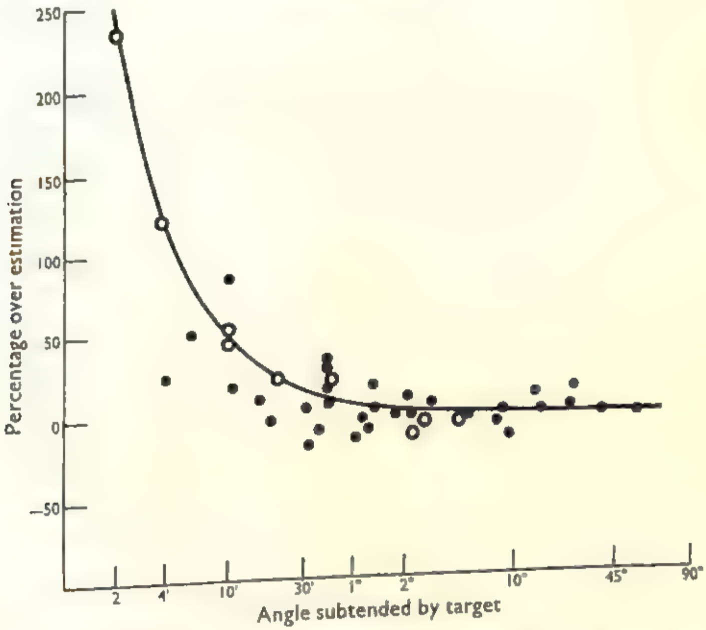
It is a physical consequence of blur that the shapes of the transitions from white image to black background are practically independent of bar width, up to the widest ones;

that is, if the images are slid sideways until the edges which caused them are made to coincide, the sigmoid curves denoting the drops in image illuminances will also coincide for all but the smallest bars.

The presence of indistinct boundaries on a perceived object poses a difficulty to someone trying to estimate its size. Is the object's edge assumed to be wherever retinal image illuminance exceeds 50 per cent. of its maximum? Or is the criterion closer to 5 per cent.? If one can assume this kind of a "threshold" criterion at all, it is immediately seen that blur will add a *fixed amount* to the perceived size of a target as a consequence of the regularity of the curves in Figure 1. Stated differently, the percentage over-estimation in judging size from a blurred image will decrease with increasing size of the object. An infinitesimally small object, as a star, generates a finite image on the retina and will have its size infinitely overestimated. An object greater than one or two degrees in extent is relatively unaffected by blur and will be estimated as accurately as the judgement criterion permits (Epstein, Park, and Casey, 1961).

The final results graphed by Johnson, Newson and May agree perfectly with the present assumption that the estimated size of a small target is too large by an amount in fixed proportion to the optical blur present. By this hypothesis the relation between overestimation and true size is hyperbolic; indeed, when the target's true size is divided into 5 minutes of arc their final points are fitted well, as shown in Figure 2. The open

FIGURE 2



The data of Joynson, Newson and May (1965) fitted (solid line) by the function 5′/(target size) predicted from optical blur.

circles denote significance at 0·05 or below, based upon a t test. The solid curve represents $5'/x$, where x is the angle subtended by the target. The quantity, 5 minutes of arc, corresponds to the amount added to a target's width by the observer's threshold response to blur. The author obtained the value 5 minutes by choosing it so that the resultant hyperbolic curve fit Johnson's *et al.* data at one point; that it fit at all other points as well supports the present hypothesis. This author's calculations of retinal image illuminances for a 4 mm. diameter pupil, and the choice of 5 minutes as the arbitrary constant, imply that the observers regard a target's "edge" as existing where its image illuminance rises about 8 per cent. toward its maximum value.

Many experiments on size constancy have been performed with test targets subtending visual angles less than one degree, where optical blur is not negligible; it is when measurements with such small targets are attempted that inherent imperfections in the eye's

optics enter into the results. The concluding figure of Joynson, Newson and May appears to be a vivid illustration of the point.

References

ARNULF, A., and DUPUY, O. (1960). La transmission des contrastes par le system optique de l'oeil et les seuils de contrastes retiniens. *C.R. Acad. Sci., Paris*, **250**, 2757–9.

CAMPBELL, F. W., and GREEN, D. G. (1965). Optical and retinal factors affecting visual resolution. *J. Physiol.*, **181**, 576–93.

CAMPBELL, F. W., and GUBISCH, R. W. (1966). Optical quality of the human eye. *J. Physiol. (Lond.)*, **186**, 558–78.

EPSTEIN, W., PARK, J., and CASEY, A. (1961). The current status of the size-distance hypotheses. *Psychol. Bull.*, **58**, 491–514.

JOYNSON, R. B., NEWSON, L. J., and MAY, D. S. (1965). The limits of over-constancy. *Quart. J. exp. Psychol.*, **17**, 209–16.

Manuscript received 15th February, 1966.

"SLOPING WATER" AND RELATED FRAMEWORK ILLUSIONS: SOME INFORMAL OBSERVATIONS

BY

N. F. DIXON and P. M. DIXON

From the Department of Psychology, University College, London

A series of uncontrived experiences, of a somewhat unusual kind, encountered while sailing, have brought to light some framework illusions which may not be familiar to most people.

The "experimental" situation is that of being in the cabin of a deep keeled yacht which, having grounded, begins to list with the receding tide. An added and interesting feature of the situation is that as it lists the yacht slowly fills with water.

Observations over periods of up to 12 hr. made by two and in some cases three subjects may be summarized as follows:—

(*a*) The increasing phenomenal visual list or movement out of the vertical of normally vertical components of the yacht's interior lags behind the real list as indicated by an inclinometer fixed to the bulkhead.

(*b*) Reference to the inclinometer while possibly occasioning anxiety does not increase the phenomenal visual list.

(*c*) Knowledge that the inclinometer is a simple and reliable instrument (a weighted arm pivoted above a quadrantal scale) does nothing to reduce the illusion that it is the pointer itself which is moving further and further out of the vertical, i.e. approximately 30 per cent. of the list is phenomenally imputed to one of the only true verticals left in the immediate environment.

(*d*) When a state of equilibrium has been reached with a constant list of anything from 10° to 45° out of the vertical, there occurs a progressively decreased phenomenal list. Things maintaining a true vertical such as hanging coats and other people appear markedly displaced from the vertical (Koffka, 1935). It is possible that the Honi phenomenon (Wittreich, 1959) operates here in as much as people known to one appear more vertical than dogs and other objects.

(*e*) Moving about in this environment is more difficult than it should be, due presumably to the misleading visual frame. It is interesting that postural cues both gravitational and kinaesthetic do not, even when one is moving about, overcome the illusion that actually sloping surfaces appear more vertical or horizontal than they should be.

(*f*) Most striking of the various illusions encountered is the appearance of the water which has entered the cabin. With its surface extending from the cabin floor on one side of the yacht to deck level on the other, it appears to be steeply slanted upwards. So powerful is this effect that observers expressed surprise at the fact that a match dropped on the water does not immediately slide towards the floor. It appears as if glued to the surface. Deliberate consideration of the fact that water inside a stranded yacht, like water anywhere else, finds its own level does nothing to alter this illusion. It is a convincing example of the prepotency of the phenomenal vertical over knowledge of the true horizontal.

(*g*) After a period of 6 to 10 hr. the real environment returns once more to an "even keel." Despite the fact that the total time spent in a tilted environment exceeds that of the time during which list is decreasing, there are no discernible after effects. Verticals appear vertical and not tilted in the opposite direction.

Discussion

Judging from their observations, the subjects in this "experiment" might be regarded as highly "field dependent" (Witkin *et al.*, 1954) in that their judgements of verticality and horizontality were primarily determined by the visual frame. It is perhaps significant that the situation is one which invites a feeling of helpless passivity, a waiting for events (the flood tide) over which the subject has no control. This state of affairs is consistent

with the relationship found by the aforementioned workers between personality factors and visual dependency.

It is possible that phenomenal restitution of the vertical is to some extent motivated by anxiety. However, no difference in the strength of the illusion was noticed as between the first and subsequent occasions upon which it was observed, despite the fact that the first occasion was considerably more disturbing than those thereafter. How a boat, stuck in the mud and half full of water, would react to a returning tide is a question productive of depressing speculations until the event has been experienced. Again it appears to make little difference as to whether the experience occurs in sheltered waters or on a shoal in open sea, despite an appreciation in the latter case that a worsening of the weather can cause a yacht to break up before floating.

Possible support for the view that slow listing out of the vertical is not, *per se*, productive of anxiety comes from observation of a large dog of mixed parentage who witnessed each occurrence. On the occasions described, this subject manifested extreme anxiety but, in a number of "control experiments" carried out in a subsequent and more robust vessel which did not fill with water when listing, the same animal remained calm and self-possessed.

We would tentatively conclude that as for Pavlov's dogs it is the sight of large quantities of water in the wrong place which is really productive of anxiety and not a discrepancy between visual and postural frames. Whether the anomalous experience of apparently *sloping* water adds to this anxiety is a matter for further research.

One further point of interest is that these illusions imply determination by generalized and long standing, as opposed to specific, schemata. Unlike a room, as used in the Witkin experiments, the cabin of a small sailing boat is very rarely vertical. If past experience of this *particular* environment determines the illusion, it can only be because the observer uses as his yardstick the algebraic mean of all inclinations hitherto experienced. Whatever the origins of the schematic determination in this situation the problem remains of reconciling our observations with data from studies involving aniseikonic lenses. In the latter experimental situation, it has been reported (Ames, 1946), that, while the wearer of aniseikonic lenses may see a horizontal lawn tipped up, static water appears *level* and further away. In discussing this phenomenon Ames remarks that any explanation must be based on the fact that, "the perceiver knows from experience that grass surfaces can be inclined but that the surfaces of bodies of water are always horizontal."

Evidently the interior of a tilted yacht confers an illusory impression not yielded by aniseikonic vision.

References

AMES, A. (1946). Binocular vision as affected by relations between unioclar stimulus-patterns in commonplace environments. *Amer. J. Psychol.*, **59**, 333–57.

KOFFKA, K. (1935). *Principles of Gestalt Psychology*. London: Routledge & Kegan Paul.

WITKIN, H. A., LEWIS, H. B., HERTZMAN, M., MACHOVER, K., MEISSNER, P. B., and WAPNER, S. (1954). *Personality Through Perception: An Experimental and Clinical Study*. New York: Harper.

WITTREICH, W. J. (1959). Visual perception and personality. *Scient. Amer.*, **200**, 56–75.

Manuscript received 10th March, 1966.

A COMPARISON OF TWO FLASH AND TWO CLICK THRESHOLDS IN SCHIZOPHRENIC AND NORMAL SUBJECTS

BY

P. H. VENABLES

From the Department of Psychology, Birkbeck College, University of London

The thresholds of fusion of paired flashes and paired clicks were compared. It was found that schizophrenics tended to show relatively higher click thresholds than normal subjects and in the most deteriorated schizophrenics two click thresholds tended to be higher than two flash thresholds.

Introduction

Wells and Kelley (1922) reported that dementia praecox patients showed a *relatively* longer reaction time to auditory than to visual stimuli as compared to a group of normal subjects. This finding was replicated by Venables and O'Connor (1959*a*) who showed that in the most ill patients, reaction time to sound was *absolutely* longer than to light. Whether or not there is an inversion of the usual modality order in reaction time is undoubtedly a function of the relative intensity of the stimuli in the two modalities and additional evidence for a relative deterioration in efficiency of functions involving the auditory modality is required.

The threshold of fusion of paired flashes which may tentatively be thought of as a measure of cortical reactivity has provided reliable results in work with schizophrenics (Venables and Wing, 1962; Venables, 1963). Furthermore, unpublished work has shown that in contrast to critical flicker fusion threshold there is no significant alteration in two flash threshold over a range of flash intensities from 50 to 2,000 foot-lamberts. It was therefore decided to carry out an experiment comparing fusion thresholds for paired stimuli in the auditory and visual modalities in order to find whether the results on reaction time could be repeated with another form of task.

Method

Apparatus and procedure

Paired stimuli, either flashes or clicks, were produced by pairs of pulses, 0·1 millisec. in duration.

The inter-pulse interval could be varied between 2 and 155 millisec. to the nearest millisec. For the production of flashes the pulses were applied to the grid of a Ferranti CL 66 cathode ray tube. This tube has a white phosphor with a decay time of 5 μsec. The flash intensity used was approximately 250 foot-lamberts. The light source had a visual angle of 1·8° at a viewing distance of 4 ft. Flashes were viewed binocularly. A fixation patch of 0·125 foot-lamberts intensity was provided by the edge illumination of a perspex block placed in front of the tube face.

Clicks were produced monaurally by applying the 0·1 millisec. pulses across an ear-piece of a pair of moving coil headphones. Preliminary experiments suggested click-fusion threshold was independent of intensity provided the click intensity was not large. A pulse of 10 volts amplitude was used in the work reported, this gave a click which was approximately 60 db. above threshold. Owing to the difficulty of measuring click intensities it is not easy to be accurate in statements of magnitude for these sounds.

Testing was carried out in a quiet, completely darkened room. The subject reported whether he perceived one or two stimuli by pressing a button once or twice. Threshold of fusion was measured by the modified form of the constant method described in Horn and Venables (1964). Thresholds were measured in the following order: (1) clicks to left ear; (2) clicks to right ear; (3) flashes; (4) flashes; (5) clicks to right ear; (6) clicks to left ear.

Subjects

Twenty normal subjects, 11 men and nine women, were tested. They were members of the artisan nursing, medical and auxillary staffs of the hospital in which the experiment was carried out. The mean age of the normal subjects was 31·9, SD 10·9, years.

Forty schizophrenic subjects were tested but five were too disturbed to enable satisfactory thresholds of fusion to be determined. The mean age of the remaining 35 subjects of whom 25 were men and 10 women was 47·3, SD 8·9, years, they were thus older ($p < 0{\cdot}01$) than the normal subjects, but were all under 58 years of age. They had all been in hospital for more than 2 years and could therefore be considered chronic patients, their mean stay in hospital was 14·1, SD 5·8, years. No patient was receiving tranquilizing drugs at the time of the experiment or for 1 month before, or was on night sedation. No patient was leucotomized or had any history of organic cerebral disease or epilepsy. All patients at some time during their illness had a history of typical schizophrenic symptoms such as thought disorder, delusions, or catatonic motor disturbance.

The schizophrenic group was sub-divided into five groups on the basis of a measure of "deterioration." This was measured by items 1 and 5 of the scale described by Venables and O'Connor (1959*b*). Each patient was rated by two nurses and the inter-rater reliability of the measure was 0·88.

Results

Two measures were available for each threshold, and reliabilities were therefore calculated. They were, for two flash thresholds; 0·93 for the schizophrenic data and 0·93 for the normal data. For mean two click threshold, reliabilities were, 0·91 for the schizophrenic data and 0·94 for the normal data. The click thresholds were the means for the two ears, while differences between ears are the subject of a later paper. The data used in later analyses were the means of two measures used above in the calculation of reliabilities. The mean two flash threshold was 83·8, SD 27·8, millisec. for the schizophrenics and 85·6, SD 30·2, millisec. for the normal subjects. The mean two click threshold was 52·5, SD 42·6, millisec. for the schizophrenics and 34·0, SD 30·5, for the normal subjects.

The groups do not differ significantly on either of these thresholds. The differences between thresholds for the auditory and visual modality was calculated for each subject and each group. For schizophrenics the mean difference was +31·3, SD 33·1, millisec. with a range from +91 to −38 millisec. A positive figure indicates that two flash is higher than two click thresholds. For normal subjects the mean threshold difference was +51·6, SD 21·9, millisec. range +94 to +11 millisec. The difference between the threshold differences for the normals and schizophrenics is significant ($t = 2{\cdot}47$, $p < 0{\cdot}02$). Within the schizophrenic group, analysis of variance between schizophrenic sub-groups showed that deterioration had a significant effect ($F = 4{\cdot}14$, $p < 0{\cdot}01$, $d.f.\ 4/30$). The least deteriorated schizophrenics had a mean threshold difference of +50·4 millisec. which did not differ from that of the normal subjects. The most deteriorated patients had a mean threshold difference of −34 millisec. The relation between flash-click threshold difference and deterioration was linear and could be expressed as a correlation of −0·48 ($p < 0{\cdot}01$). As it seemed likely that deterioration might be related to age and length of stay in hospital, correlations between it and these variables were calculated. The correlation between age and deterioration was +0·21 which was not significant. However, that between length of stay and deterioration was +0·43 ($p < 0{\cdot}01$). In view of this significant correlation the correlation between length of stay and flash-click threshold difference was calculated. It was found to be −0·06. The partial correlation between deterioration and flash-click threshold differences with length of stay held constant was found to be −0·48 ($p < 0{\cdot}01$). In view of the difference in ages of the normal and schizophrenic groups the correlation between age and flash-click threshold differences was calculated, and it was found to be 0·18 (NS). Age may thus be presumed to be an unimportant factor in the results reported, where no subject was over 58 years old.

Discussion

The results show that schizophrenic pathology appears to involve a relative deficiency in performance involving the auditory modality.

Other data are available which support this contention to a greater or lesser extent. Zahn, Rosenthal and Lawlor (1962) described work on the GSR orienting reactions to auditory and visual stimuli in normal and schizophrenic groups. Examination of the

figures which they presented shows that the basal log conductance level during the series of stimuli was greater with visual than with auditory stimuli in the schizophrenics and that the reverse was the case with normals.

A similar pattern with reaction times to that found by Venables and O'Connor (1959a) was shown in a study by Sutton, Hakerem, Zubin and Portnoy (1961) in which there was a general tendency for RTs to be faster to light than to sound in chronic schizophrenic patients.

This study and others which were reviewed by Sutton and Zubin (1963) all show that, with schizophrenic subjects, reactions to sound stimuli are impaired if the previous trial was a light, whereas reactions to visual stimuli however preceded do not yield differences between patients and normals.

Spain (1964) in a study of eyelid conditioning in schizophrenic and normals has shown better conditioning to a visual CS than to an auditory CS in schizophrenics and the reverse in normal subjects.

A final suggestive piece of evidence is that of Abramson, Jarvik and Hirsch (1955) who show that simple RT to sound is impaired by the use of LSD-25 while simple RT to light is not.

Further investigation of relative modality impairment on men and animals should lead to better understanding of the schizophrenic process.

This work was carried out at Netherne Hospital, Cousdon, Surrey, by kind permission of Dr. R. K. Freudenberg.

References

Abramson, H. A., Jarvik, M. E., and Hirsch, M. W. (1955). LSD 25 X. Effect on RT auditory and visual stimuli. *J. Psychol.*, **40**, 39–52.

Horn, G., and Venables, P. H. (1964). The effect of somaesthetic and acoustic stimuli on the threshold of fusion of paired light flashes in human subjects. *Quart. J. exp. Psychol.*, **16**, 289–96.

Spain, B. (1964). *Parameters of Eyelid Conditioning in Normal and Schizophrenic Subjects.* Unpublished M.A. thesis, University of London.

Sutton, S., Hakerem, G., Zubin, J., and Portnoy, M. (1961). The effect of shift of sensory modality on serial reaction time: a comparison of schizophrenics and normals. *Amer. J. Psychol.*, **74**, 224–32.

Sutton, S., and Zubin, J. (1965). Effect of sequence on reaction time in schizophrenia. In Welford, A. T., and Birren, J. E. (Eds.), *Behaviour, Aging and the Nervous System*, Chap. 27. Springfield: Thomas.

Venables, P. H. (1963). The relationship between level of skin potential and fusion of paired light flashes in schizophrenic and normal subjects. *J. Psychiat. Res.*, **1**, 279–87.

Venables, P. H., and O'Connor, N. (1959a). Reaction time to auditory and visual stimulation in schizophrenic and normal subjects. *Quart. J. exp. Psychol.*, **11**, 175–9.

Venables, P. H., and O'Connor, N. (1959b). A short scale for rating paranoid schizophrenia. *J. Ment. Sci.*, **105**, 815–8.

Venables, P. H., and Wing, J. K. (1962). Level of arousal and the sub-classification of schizophrenia. *Arch. gen. Psychiat.*, **7**, 114–9.

Wells, F. L., and Kelley, C. M. (1922). The simple reaction in psychosis. *Amer. J. Psychiat.*, **2**, 53–9.

Zahn, T. P., Rosenthal, D., and Lawlor, W. G. (1962). G.S.R. orienting reactions to visual and auditory stimuli in chronic schizophrenic and normal subjects. *Paper presented to the Society for Psychophysiological Research, Denver, Colorado.*

Manuscript received 21st May, 1966.

APPARATUS NOTE

A SIMPLE AID FOR RECORDER CHART ANALYSIS

BY

J. A. M. HOWE

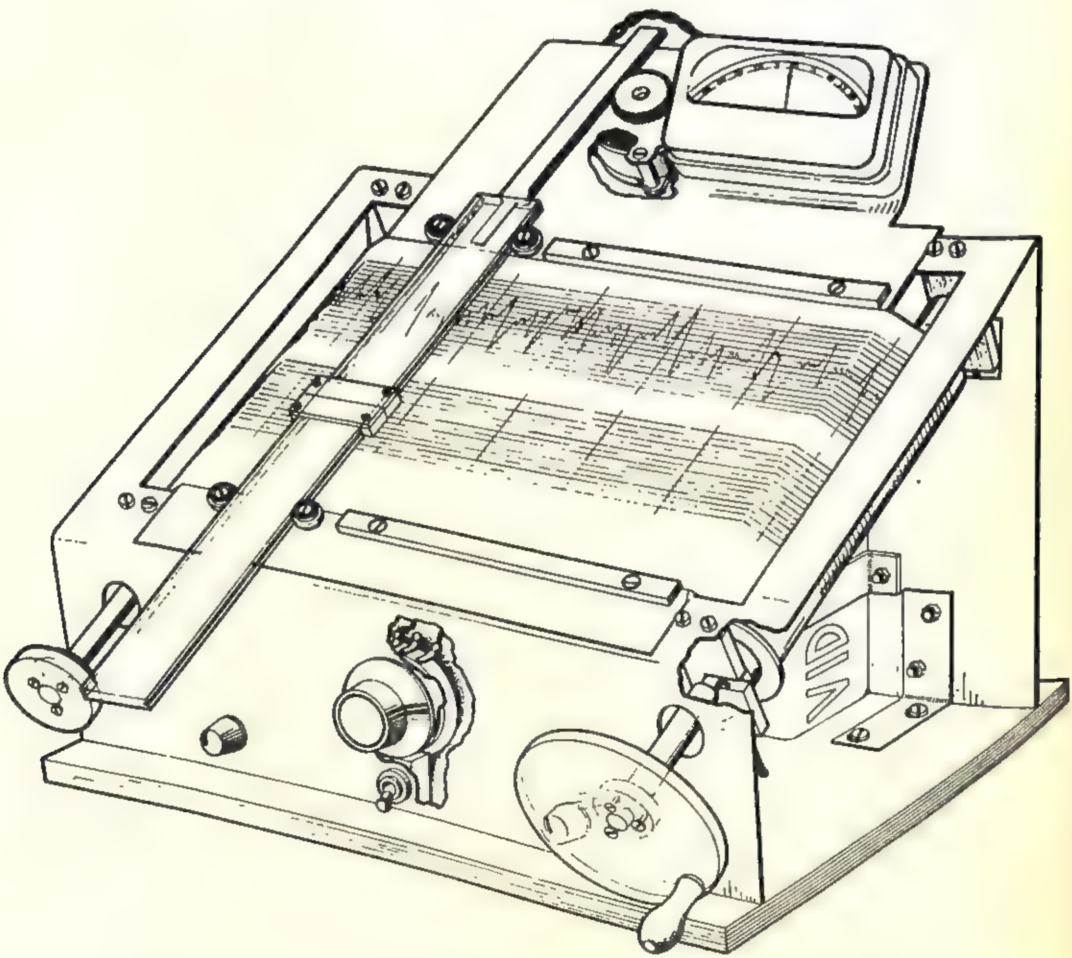
From the Psychological Laboratory, Cambridge

Purpose

The data from many psychological experiments are recorded as displacements of a pen across a moving paper strip. Quantification of the information contained is a time-consuming process and is extremely tedious when a large number of measurements have to be taken. Inaccuracy in measurement may result; research progress is delayed, and the thought of having to carry out such analysis may deter an experimenter from pursuing a particular line of research. The spread of computing techniques may eventually reduce this problem to manageable proportions, but the necessary apparatus is at present expensive and cannot be directly applied to the processing of certain data, e.g. the record of eye position obtained by electro-oculography.

To facilitate the scoring of a large number of chart records of eye movements, the apparatus shown in Figure I was constructed. It is inexpensive and can be assembled by a competent technical assistant.

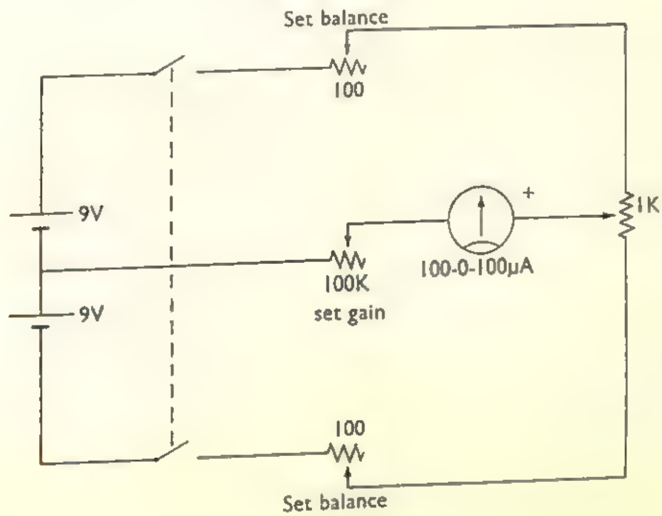
FIGURE I



Design

The instrument chassis consists of a sheet of 16 gauge dural plate, folded to provide a working surface at 30° to the horizontal and mounted on a baseboard. A detachable roller is mounted on the under side of the plate at each end, slots being cut in the plate to allow access of the chart to the rollers. A perspex bar is located between ball races screwed to the surface of the plate and is free to slide up and down. The ball races are fitted with rims which engage with slots cut in either edge of the perspex bar. The cursor can be moved freely to any position, and a metal rack attached to the upper end of the bar engages with a pinion which is secured to the shaft of a precision potentiometer. This is connected to the meter, which is calibrated in suitable units. The circuit diagram, Figure 2, indicates the connection of the components, the remaining ones being mounted on the vertical front plate of the apparatus, the "balance" potentiometers flanking the "set gain" potentiometer and the on-off switch.

FIGURE 2



Method of operation

A chart is prepared by interpolating the zero reference lines for all trials. It is then wound on to the right-hand roller of the machine and fed across beneath the perspex bar to the left-hand roller. If the charts have more than one channel of information the components may be lined off across the chart width as it is wound through to synchronize the position at which measurements are made (in many cases the whole of this preparatory procedure would not be required). The perspex bar is located in a centred position by means of a spring-loaded ball catch, and differential ageing of the supply batteries can be compensated by adjusting the balance controls to zero the meter. The cursor is then pushed along the bar until the hair-line on it corresponds with the zero of a calibration trace, recorded on the chart by introducing a *known* signal at some time(s) during the experiment. By moving the bar up or down from this zero until the hair-line is adjacent to the displacement of the trace produced by this known signal the gain of the electrical part of the system can be adjusted by means of the set-gain control until the meter reading corresponds to the value of the calibration displacement. Movement of the bar rotates by means of a high precision anti-backlash gearing the 1K potentiometer which gives an electrical analogue of the co-ordinate.

The system is thus set for measuring unknown displacements, and an important design feature of the instrument is that zero drift can be accommodated by returning the perspex bar to the centred position (zero on the meter), and moving the cursor to keep the hair-line located on the interpolated zero reference line. This process does not alter the overall gain of the system, but just acts in a way similar to the shift control on an oscilloscope. The positive and negative displacements from zero can be read off directly from the meter scale, eliminating conversion calculations in interpreting the records, and greatly shortening analysis time. The accuracy of the apparatus is dependent

upon the precision of the rack and pinion gear assembly, and the resolution and linearity of the potentiometer and meter. (By replacing the meter with a digital voltmeter, errors in reading can be eliminated and resolution can be increased.) However, given that good quality components are used, the final resolution will depend upon the thickness of the trace.

This apparatus can be used to process the data from many different types of experiment. It is not restricted to the analysis of eye fixations although it was specifically developed for this purpose. It can be used with charts of different widths, and the main limitation is that only y-axis measurements can be made. However, it would be possible to construct a second machine with the perspex bar along the chart for x-axis measurements.

This device was constructed while the writer was supported by a D.S.I.R. Studentship.

Manuscript received 16th March, 1966.

BOOK REVIEWS

New Horizons in Psychology. Edited by Brian M. Foss. London: Penguin Books. 1966. Pp. 446. 7s. 6d.

When 21 authors contribute short chapters to a popular book on what's new in psychology the result may too easily be unco-ordinated, repetitious and uneven in content and style. Brian Foss's book is none of these. With considerable editorial skill he has brought together a team of British psychologists whose papers on subjects ranging from social psychology to electrophysiology provide an absorbing "action picture" of modern psychology.

The scope of the book is huge. Separate chapters cover learning theory, behavioural genetics, invertebrate behaviour, selective attention, creativity testing, small groups, cross-cultural studies, behaviour therapy—and much more. The book is divided into five sections containing related topics and this together with a well written commentary by the editor gives it a reasonably systematic plan. The chapters are similar in their design, each beginning with a survey of past work in the field and taking up a more detailed discussion of recent experiments and ideas.

Most authors have wisely avoided trying to be exhaustive and have chosen to discuss a few aspects of their subject at fair length often with particular emphasis on their own work. Gregory for instance in his chapter on visual illusions considers early theories of the illusions and goes on to talk about his own version of the "perspective theory" in the light of new evidence from luminous illusion figures. Wason's chapter on reasoning consists largely of an account of two fascinating experiments of his own on the derivation of rules in abstract problem solving. Dixon discusses subception, Wilkinson his work on sleep deprivation, Joyce his studies of the clinical trials of drugs. Almost all the authors present interesting new material in a setting which enables a relatively uninformed reader to appreciate its sense and significance.

Much of the material here is unavailable outside journals or specialist monographs and some of it is unpublished. The book may therefore be of special value to students who have not the time or opportunity to keep up with new developments. The index is good and the references are extensive. Several chapters are in a way introductory primers to their respective branches of psychology, notably Brown's on information theory, Dalrymple-Alford's on psycholinguistics, Connolly's on the genetics of behaviour and Stretch's on operant conditioning, and for educational purposes these should be doubly useful. The choice of topics is too idiosyncratic to make the book suitable as a general text but as an accessory reader it should add some spice to the all too bland flavour of the established instructional texts.

The book is cheap and will be widely distributed. It may therefore have considerable influence on the public image of psychology in Britain and it is important to know what kind of picture emerges. Apart perhaps from Bannister's chapter on personal construct theory, psychology is presented as a strictly experimental science. The tenor is anti-theoretical in the sense that the emphasis is on controlled observation in defiance of dogma. In one chapter after another a theoretical position is stated only to be undermined or amended in the light of recent experimental work. Thus Treisman in her chapter on selective attention shows up the limitations of filter theory in its original form, Brown discusses the difficulties involved in straightforward application of information theory to human performance, Tyson is cautious in her assessment of the value of creativity tests, Sluckin considers theories of the effects of maternal deprivation to need revision, Annett suggests that the superiority of teaching machines to teachers has not been clearly demonstrated, Beech emphasizes that behaviour therapy is still very much at a formative stage.

Yet cautious as it is the book is not negative. The impression it gives is of psychology as a young science which having overreached itself in making the sweeping generalizations of the immediate post-war years, has lately been taken to task for its precocity. The optimism of earlier theorists was premature. But their theories served well in stimulating the research which eventually required their own revision and while it may seem that theoretical psychology has taken a step backwards in the last few years there is every reason to suppose it is not a retreat but a case of *reculer pour mieux sauter*.

N. K. HUMPHREY.

Eye and Brain: The Psychology of Seeing. By R. L. Gregory. London: Weidenfeld and Nicolson (World University Library). 1966. Pp. 254. 14s.

This book encompasses a number of topics classically subsumed under the classifications of visual perception, physics of light, neurophysiology, art, and zoology. Although at first glance it might appear that such diversity could not logically be treated under the same heading, the author shows, expertly and delightfully, how a consideration of the basic process of seeing must, in fact, encompass a wide variety of topics. Indeed, Gregory's approach has been to ignore traditional classification schemes and to delve into the process of seeing without regard for the usual artificial distinctions between sensation and perception, zoology and psychology, physics and physiology. The result is delightful. Gregory treats, in a most readable style, the physics of light, anatomy and physiology of the visual system, evolution of the eye, brightness, motion and colour perception, visual illusions, art, perceptual learning, and some problems of vision in space, with an imaginative and eclectic approach.

The initial impression made by this book is the high quality of the skilfully prepared and elegantly presented art work. The text contains more than 100 figures, diagrams, and photographs, some of which are printed in high quality colour. For the first time the reader of the visual literature is treated to a text in which an attempt is made to utilize the full capacity of the human visual system, including colour and stereo, in the elucidation of the visual process itself. A number of the illustrations are original and make an impressive contribution to the literature in addition to their value in clarifying and amplifying the text material.

Several chapters merit special attention. The chapter on visual illusions is one of the most comprehensive and original treatments of this ancient topic ever published. The theoretical position that illusions represent the misapplication of normal perceptual mechanisms, which has been discussed briefly elsewhere, is here treated in depth along with some original experiments and thoughts by Gregory. The subsequent chapter entitled "Art and Reality" expands the treatment to include three-dimensional reproductions which could, logically, be considered along with the problem of illusions. Taken together, they represent a noteworthy and original contribution which would be inappropriate in today's space starved journals but which fit nicely into the atmosphere created by this book. The chapter entitled "Do We Have To Learn To See?" treats the problem comprehensively and historically, and includes the recent research by Gregory and Wallace on a man whose sight was restored after 52 years of blindness. This research and discussion will do much to offset some of the misconceptions ingrained in modern psychology by the work of von Senden. The section on "Seeing Movement" also presents some of the author's original research while treating a number of general problems of motion perception. The chapter on "Seeing Colour" shares the clear exposition and illustration of the entire book, but treats only a few elementary concepts.

Particularly impressive to this reviewer is the awareness which Gregory exhibits regarding psychology's historical background and its logical continuity with philosophy. The book conveys the impression that visual perception is essentially a topic which has interested individuals in many disciplines throughout recorded history. Unfortunately, most treatments ignore the historical antecedents which are necessary for a meaningful understanding of the present state of the art. Gregory's familiarity with philosophy, history, anatomy and physiology, as well as his knowledge of the Gestalt literature and its antecedents, all contribute to the flavour and value of this book.

The reader should be warned, however, that this is *not* a textbook in vision or perception. Some topics which usually appear in such textbooks are not mentioned at all, while others are given special treatment depending, no doubt, on the particular interests of the author. The experienced reader will notice some unevenness of treatment. A number of statements are made for which no references are given. These comments are not meant as a criticism in view of the objectives of the book and the excellence with which the included subject matter is handled. Rather, they are mentioned to avoid disappointment on the part of those who expect a formal textbook.

In summary, this book will stimulate the imagination and add to the knowledge of readers at all levels of experience irrespective of their backgrounds. In fact, this reviewer would take the position that anyone reading this book will not only find it to be a delightful intellectual experience, but will feel that his personal knowledge of man's most important sense has been significantly incremented.

HERSCHEL W. LEIBOWITZ.

Vision and Visual Perception. Edited by Clarence H. Graham. London and New York: Wiley. 1965. Pp. vii + 637. £9.

This book is in no way novel or startling. It aims at providing "a useful listing of topics which should be discussed in a general coverage of vision," and "to stand as a point of departure for an understanding of ensuing developments." The phrase "a general coverage of vision" is perhaps a little misleading for a book which does not mention developmental, learning, clinical or, with a few exceptions, animal behavioural studies. The point of view is that of the psychologist concerned with the normal functioning of sensory and perceptual mechanisms in the adult human.

The chapters are: Riggs on Light as a stimulus for vision, Electrophysiology, and Visual acuity; John Lott Brown on the Structure of the visual system, Brightness discrimination and contrast (with Mueller), Flicker, Colour contrast and constancy (with Graham), and After images; N. R. Bartlett on Absolute thresholds and on Dark and light adaptation; Hsia on Photochemistry and on Colour blindness (with Graham). The editor, C. H. Graham, contributes two introductory chapters, later ones on Discriminations that depend on wavelength, Colour mixture and colour systems, Colour theories, and three final chapters on Space, Form and Movement perception.

This book will be useful to researchers, teachers and advanced students. The chapter on acuity is a beautifully organized summary of a confusing field. It is good to have extended treatments of after images and of flicker, including some of the recent work based on the Fourier analysis of the stimulus. Almost every chapter contains some unfamiliar references, but often the coverage is not quite as up-to-date as one could wish, doubtless because the book has taken 12 years to produce. There are some curious omissions. For example, there is no mention of the autokinetic effect, the spiral (or waterfall) after-effect, binocular rivalry, artificial di- or monochromatism after adaptation to bright lights, or small-field dichromatism. There are few references to Russian work.

The various topics are in general treated in the traditional way. The authors have done well what they set out to do, and therefore this reviewer feels he is being a little ungrateful in stating, as he must, that he was repeatedly disappointed that they had not set out to do more. Although this is a scholarly book, very well documented, it is also somewhat uncritical, or at least reluctant to evaluate. It abounds in phrases like "it is too early to make anything like a final evaluation of. . . ." Of course this is often appropriate, but it does sometimes seem to be an easy way out. Often the treatment seems too conventional. Much of the material was already available in book form, for example in the writings of Le Grand, Granit, Brindley, and Graham himself. The pre-existing accounts remain in many ways superior. Le Grand's text is still unsurpassed in clarity; Brindley's book still contains the most trenchant arguments. It is of course easy to complain that the treatment is too conventional. What could have been done instead? It is chiefly that many of the categories into which the subject matter is divided are by now somewhat stale. Some new ones could have been formed. Three examples come to mind of topics on which easily available discussions are badly needed, and which would have fitted into the scope of this book. First, a discussion of the various types of visual masking. Secondly, a comparison and discussion of the measures of "integration time," "latent period," etc., of the eye and the visual pathway, which arise from studies of Bloch's law, the Pulfrich phenomenon, visual masking, flicker, reaction times and physiological work. Thirdly, a little further from the scope of the book, a comparison of visual and auditory data and mechanisms. Many other examples will occur to anyone familiar with the literature on vision. This failure to connect facts in novel ways is reflected in the organization of the text as lists of experiments within a given field, or in the theoretical sections under a series of headings of the form "so-and-so's theory of. . . ." The latter is not only unpleasantly reminiscent of the "schools of psychology" approach, but also tends to allow some important theoretical issues to slip through the net so that they are not discussed at all. Organizing the discussion not around theories or around particular pieces of work, but around issues, such as "Are there three and only three colour channels?" can avoid this, and also seems more scientifically mature. As the book is arranged it is often hard to find the answer to questions of this sort. Surely we try to think, in our better moments, about issues, and not about whether so-and-so was right or not?

It is perhaps part of the same syndrome that some recent attempts to provide unifying explanations are not well described. Thus in the section on the Fourier analysis of a flickering stimulus, the concept of linearity is barely mentioned. The concept of an "equivalent background" which clarifies considerably the subjects of dark adaptation

and after images, is not mentioned, although some of the experiments on which it is based, and some closely related ideas, are described.

This review has been written on the assumption that reviewers should criticize. The book does maintain a high standard of accuracy and of comprehensiveness within its chosen field. There is more in it to praise than has been indicated here. One's regret is that it does not provide more new insights. PAUL WHITTLE.

The Critical Flicker Frequency Series Effect. By Károly Ákos and Magda Ákos. Budapest: Akadémiai Kiadó. 1966. Pp. 245.

This monograph describes a technique which consists of obtaining a series of 50 successive descending judgements of CFF. The form of the fluctuations in this series is found to vary considerably between individuals but to be more or less repeatable for any one subject. It is suggested that this "CFFSE" may be a useful clinical tool.

PAUL WHITTLE.

Curiosity and Exploratory Behaviour. By Harry Fowler. London: Collier-Macmillan. 1965. Pp. viii + 216. 15s.

This book is one of the first titles in a new paperback psychology series. It is intended for undergraduate use and is divided into two sections; the first third of the book is a review of the current state of information on exploratory behaviour, while the remainder is devoted to a selection of original papers.

Fowler's review section begins with a rather general discussion of the nature of explanation in psychology which could have been omitted with advantage. Most undergraduates are over-exposed to this sort of argument, and in this case the exposition is neither satisfactory nor elegantly phrased. The remainder of the review is, however, thorough and thoughtful. Fowler begins by considering the basic empirical facts about exploratory behaviour and then develops the "empirical tenet" that all exploratory behaviour can be treated as a response to changes in the complex of stimulation. He then goes on to theories of exploration, distinguishing between those based on curiosity and those based on boredom. The attempt is then made to fit both of these types of theory into something like a Hullian framework, with boredom functioning like an ordinary appetitive drive, increasing with deprivation, while curiosity is developed as an incentive motivation on the basis of fractional anticipatory goal responses. The attempt is ingenious and interesting, but it is not, I think successful; it fails in particular over the lack of empirical support for the boredom drive and the difficulty in specifying a credible goal response. Considerable space is also devoted to the discussion of optimal level of stimulation accounts of exploratory behaviour, in particular those of Berlyne and Fiske and Maddi; the discussion and criticism is illuminating and clear. Indeed the review, as a whole, treats the subject at a level well above that of most such introductory paperbacks and can be strongly recommended to undergraduates, or indeed anyone curious about curiosity.

The selection of readings is careful and catholic. In such a case one is always aware of what has been omitted; but, although there are a number of papers which I would like to have seen included, this would have meant omitting some of the present selection, and I do not think that the result would have been an improvement. There is, however, a further point. The reprinted papers must account for a large proportion of the cost of the book, and their presence therefore requires justification. The usual pretext is that the original papers are widely scattered or largely inaccessible. But all but one of the papers in this collection were originally published in *The Journal of Comparative and Physiological Psychology* or *The Psychological Review*, and all but two since 1953. It seems to me that there is very little excuse for collecting these papers in this way when they must be accessible to any undergraduate. M. S. HALLIDAY.

Psychiatrie Animale. Edited by A. Brion and H. Ey. Paris: Desclée de Brouwer. 1964. Pp. 605. 480 FB.

The fundamental question this weighty tome—over 600 pages, over 37 contributors—poses is—can there be such a thing as animal psychiatry? In pursuit of the answer we are presented with a vast amount of information, some of it relevant and some, it must be admitted, irrelevant, on such diverse topics as the origins of ethology, the metabolic bases of sleep, the effects of cerebral lesions, bird migration and the problems of training circus animals. The contributions vary vastly in quality, some being a mere page or two,

usually without references, and others scholarly reviews with useful bibliographies. Notable among these are Chertok's account of animal hypnosis, marred though it is by a too-ready reliance on psychoanalytic concepts, Delaveleye's chapter on bird migration (though both these suffer from being less up-to-date than one would wish, since the bibliographies appear to close at 1959) and Chance's reasoned statement of his theory of the evolutionary importance of epileptic phenomena. Much less satisfying are the numerous and overtly anecdotal treatments of various aspects of allegedly abnormal behaviour in animals, and the sections on experimental neurosis, both from the Russian and American schools, are frankly disappointing. The section on behavioural inheritance appears to equate selective breeding studies with McDougall's famous Lamarckian experiment, and the chapters in the first section of the book dealing with the various viewpoints in ethology are too superficial to do more than introduce the reader to the authors concerned. An exception is Létard, Favreau and Fontaine's brilliant summary of the history of man's view of animals, and, in the same vein, Ellenberger compares a zoo to a psychiatric hospital.

But throughout much of this work we find an almost total lack of rigour: as Hinde points out in his very readable contribution on abnormal behaviour in birds, the abnormal can only be defined by reference to the normal of the species, and his conclusion that most of the examples he has been able to cite are anecdotal and, furthermore, that they probably result from dysfunction occasioned by captivity in any case, must hold for most of the work so copiously cited in this book. The answer, then, to the question this book poses must surely be "No, there can be no animal psychiatry until our understanding of the concepts of normality and abnormality in animals can at least approach that sophistication brought to bear on the problem at the human level." This book is all-too-convincing evidence that this stage has by no means been reached.

P. L. BROADHURST.

Attention, Arousal and the Orientation Reaction. By R. Lynn. International Series of Monographs in Experimental Psychology, Vol. 2. Oxford: Pergamon Press. 1966. Pp. viii + 118. 35s.

A good case can be made for the view that the most important sustained programme of research to have been conducted in psychophysiology in the last 15 years is the work of Sokolov's group in Moscow on the orienting reaction. Judging from the frequency with which this work is referred to in Western publications, its significance is at last being appreciated here and in the United States. The time-lag, however, is distressingly long: the most important publications came out in Russian in 1958 and 1959 (two symposia and Sokolov's book *Perception and the Conditioned Reflex*) and in 1961 (Vinogradova's book *The Orienting Reflex and its Neurophysiological Mechanisms*); of these, only Sokolov's book has so far been translated into English. In the (temporary?) absence of these primary sources, it is indeed true, as Professor Eysenck puts it in his Foreword, that "Dr. Lynn's book will be a God-send to all those interested in this field."

It is very much to Dr. Lynn's credit that, if and when the primary sources (on which he has obviously relied very closely) are published, *Attention, Arousal and the Orientation Reaction* will still be an extremely useful book to have on one's shelf. It gives a concise summary of the major findings the Russians have made and of many of their theories, while judicious juxtaposition of Western data, psychological and neurophysiological, both serves to fit the Russian work into a wider and more familiar framework and allows a critical evaluation of some of the claims the Russians have advanced. This makes it particularly suitable for teaching purposes. As for the research worker, though he will not find guidance on technical detail (for this, of course, the original experimental reports are essential), he will find, particularly in the chapter on "neurological models for habituation of the orientation reaction," a clear exposition of the different theories abounding in this field, and this should help him to design meaningful experiments. This last contribution is of particular value, since the various theories are based on different experimental preparations and techniques, and have not, I think, been all considered together in this way before.

The main criticisms one could level against Dr. Lynn are concerned rather with what he does not do than with what he has done. In spite of the title there is very little attempt to consider how the orientation reaction relates to attention, either in our everyday use of this word or in other experimental contexts (e.g. Broadbent's work on selective perception or the work on selective factors in animal discrimination learning). It would be easier, I think, for others to attempt to delineate these relationships if Dr. Lynn had brought out more clearly the basically *operational* definition of the orienting

responses with which the Russians are working – i.e. *any* response which is elicited by novel stimuli, irrespective of modality, and which diminishes and disappears with repeated presentation of the stimulus. In a similar vein, one might complain that the book does not sufficiently bring out the sheer importance of the topic with which it deals. Thus the evidence suggesting the operation of a cortical inhibitory system in the habituation of the orienting reaction (which Dr. Lynn presents very clearly) has far-reaching implications for the study of learning in general; unless we are to suppose that such a sensitive mechanism as this appears to be is simply not used for more complex forms of learning. Similarly, the much greater specificity of habituation of the orienting responses to the characteristics of the stimulus than is the case for conditioned responses must have important bearings on problems in the study of perception. On a more detailed level, it seems very likely that Western controversies over the nature of sensory preconditioning can be resolved by the realization that this phenomenon is a form of conditioned orienting behaviour. And, in quite another field, it seems possible from the work of Luria and Homskaya that some insight into the effects of lesions of the frontal lobes can be gained by considering this section of the cerebral cortex as being part of the system for descending cortical control of the orientation reaction.

However, these criticisms should merely serve to emphasize the need for a clear presentation of the basic facts uncovered by the Russian research; and this need Dr. Lynn has supplied extremely well. It is to be hoped that his book will be widely read.

JEFFREY A. GRAY.

Sex and Behavior. Edited by F. A. Beach. London and New York: Wiley. 1965. Pp. xvi + 592. 75s.

The past decade has witnessed a striking liberalization in attitudes towards matters relating to sex and its associated patterns of behaviour. In this connection the arts have undoubtedly played a leading role in stimulating a general awareness of the social problems involved, but the impact of scientific discovery, particularly in the realm of contraceptive procedures, must give rise to thoughts as to whether a code of ethics based on the precept that sexual and psychological compatibilities are so highly correlated as to form a satisfactory basis for permanent association between two individuals is necessarily appropriate for the society of the future. A first line of enquiry must be to determine the nature of sexual relationship within the human species for, as has been so clearly demonstrated by the work of Beach, of the late W. C. Young and of others, sexual activity comes with ascent in the phylogenetic scale, to be dominated less by endocrine and more by social factors and so to pass from the realm of innately organized to that of adaptive behaviour. The result of a concerted attack on this problem published in 1948 and 1953 by the late Alfred Kinsey and his colleagues, essentially based on enquiry into the behaviour of male and female within a restricted stratum of North American society, has been extended by Ford and Beach in their study of sexual habits in other, and less sophisticated, environments. But much time has passed since the publication of these volumes, attitudes have changed and a re-appraisal is due. The book *Sex and Behavior* might perhaps, by reason of its title and editorship, be expected to provide information on recent advances not only in this field of enquiry but also with regard to related topics such as sex differences in emotionality, motor performance, selective learning abilities, attitudes toward the opposite sex and acceptance of social status. The reader, however, whose background leads him to expect analysis of these areas of interest may well be but partially satisfied. The volume records the proceedings of two conferences, held in Berkeley, California in 1961 and 1962 and is thus tardy in making an appearance. It contains a miscellany of 22 papers dealing with subjects as diverse as the relationship of agonistic displays by birds to the evolutionary aspects of sexual behaviour, the cerebral control of the gonadotrophic functions of the pituitary gland and the regulation of sexual behaviour within a primitive human society, to cite but a few; and although an attempt has been made to assemble these contributions into some form of logical grouping it would be difficult to identify the thread which holds them together were it not for an excellent final chapter by Beach who succeeds admirably in drawing together the relevant material into a cohesive whole and in assessing the prospects for future progress in the field. One may express the hope that the multidisciplinary assembly of information provided by this book may give rise to an interdisciplinary attack on some of the outstanding problems raised.

In a short review it is not possible to refer to all contributions and invidious to select some to the exclusion of others. Of outstanding interest to those concerned principally

with the study of primate behaviour, however, will be papers by Sears and by Romney, who take a look at some of the factors leading to the sex-specification of human behaviour; by De Vore, extending the earlier observations of Zuckerman and others on the dominance-subordination relationships of captive baboons to the free-ranging situation; and the careful analysis by Harlow, based on a comparison of animals reared in the wild and in captivity, of the development of sexual behaviour in the rhesus monkey. Davenport provides a fascinating study of the regulation and ritualization of sexual relationships in a Melanesian community while, at the level of the individual, the physiology of the coital aspects of human reproduction is explored in astonishing and unexpected detail by Masters and Johnson using methods of measurement which are not revealed in their communication but which have since been fully published in book form.

These few chapters by no means exhaust the interesting material contained in this volume and another reviewer might doubtless have selected differently from topics dealing with the control of endocrine secretions and their action upon the nervous system, the environmental control of reproductive behaviour and the evolutionary and genetic aspects of the subject. The book is beautifully produced, well illustrated and at its price of 75s. good value although, as with all publications based on multidisciplinary symposia, the purchaser will provide himself with much material of but peripheral interest to him. It is excellently edited though some may agree with the writer that the inclusion of semi-verbatim reports of verbal exchanges between participants is of greater value to the participants themselves than to the world at large and could well be replaced on publication by paragraphs summarizing the salient features of their content.

J. T. EAYRS.

Structural Models: An Introduction to the Theory of Directed Graphs. By F. Harary, R. Z. Norman and D. Cartwright. London and New York: Wiley. 1965. Pp. ix + 415. 75s.

The words "structure" and "model" are to be found extensively in the literature of the social sciences, the former referring to the pattern of relationships among pairs of elements of some system, and the latter usually referring to a set of definitions and postulates concerning some aspects of the system. This book uses the term "structural model," in the beginning at least, to mean a set of definitions and axioms about an abstract system (structure) of points and lines called a linear graph, and develops theorems about these structures. The mathematical terms introduced can be given a variety of empirical meanings, but the authors are careful to mention the perils of translating the language of graph theory into that of social relations. The first requirement for such a translation is an unambiguous procedure for identifying empirical entities with the abstract elements, and only then would the deduced theorems be valid statements about the empirical world. In this sense the structural model is descriptive and the validation of such statements is a logical validation because of the assumed isomorphy between the linear graph and the empirical situation. These statements clearly have a certain suggestive potential, but this type of model becomes predictive only when additional postulates are made about the empirical system apart from the formal properties of linear graphs. The model then possesses the "virtue of fallibility" in that its adequacy can be determined empirically.

In the first chapter an axiom system for the very general theory of nets is considered and this system is restricted, firstly to define a relation, and then to define a directed graph or digraph. A digraph is defined as a finite non-empty set, V, of points together with a finite set, X, of lines, and two functions both of whose domains are X and whose ranges are contained in V, such that no line has the same first and second points, and no two distinct lines from the same point have the same second point. These two functions specify the endpoints and direction of each line, and the way in which this is done characterizes the type of digraph. The number of lines from a point is called its out-degree, written *od*, and the number of lines to a point is its indegree, written *id*. The ordered pair (od,id) is called the degree of a point ("Unlike most humanbeings, a point has an id but no superego"), and is used as a basis for classifying the points of a digraph. Then follows a discussion of the isomorphism of two digraphs, and using this concept a precise meaning is given to the notion that two empirical systems have the same structure.

The next three chapters investigate the basic concepts of distance between two points, u and v, say, and the "reachability" of v from u, and the ways in which u and v can be

"joined" together. These concepts are used to characterize the connectedness of structures and in determining a minimal set of points which can reach all the points of the digraph. Interwoven into this considerable body of definitions and theorems are many suggestions concerning the applicability of theory to empirical situations. Thus if a digraph is interpreted as a rumour network, points representing people and the line from u to v meaning "person u transmits the rumour to person v," then one can determine, for instance, whether the rumour would reach w if planted with u, or which is the smallest set of people with whom the rumour must be planted for it to reach everyone in the network.

Chapter 5 should prove very useful to many investigators because it presents a detailed discussion of the relationship between digraphs and matrices, and indicates ways of exploiting matrix algebra, and hence computers, to obtain information about digraphs. In Chapter 6 the concept of distance is used to define the compactness of a digraph and the relative centrality of its points. We are told of experiments on five-man groups in which the estimated relative centrality of a person was found to be related to his chance of being selected group leader and to his morale. An index of the relative status of an individual in a sociometric group is also defined and this index is modified to yield a measure of "prestige." The next three chapters consider the susceptibility of a digraph to the removal of some of its lines or points. In the empirical world this could correspond to studying the destruction or malfunctioning of a community when certain relationships are broken or when certain people are exterminated. In Chapter 10 there is a discussion of digraphs in which the points can be assigned numerical values, and a welcome decline in the ratio of definitions to theorems.

Chapter 11 is one of the more interesting chapters; it discusses digraphs in which there is a line between every pair of points. These digraphs are called "tournaments," and can be used to represent structures such as the preference relation revealed by the method of paired comparisons on a set of objects, or the dominance relation on a flock of chickens. A digraph is said to be transitive if it contains a line uw whenever it contains the lines uv and vw, for any distinct points u, v, w. It is often found that empirical structures of interest do not possess this property, but the concept can still be exploited if a quantitative instead of an all-or-none definition is used. Two measures of transitivity or consistency are presented and there are references to other theoretical and empirical work (as is the case throughout the book). It is then shown that if a "diabolical" subject deliberately tries to be inconsistent in the preference task, it would be difficult to distinguish his behaviour from that of one who behaves in a random fashion (the authors are perhaps too generous in regarding the latter subject as less diabolical than the former).

In Chapter 12 certain types of digraphs are discussed in which there are restrictions on the values of the out- and in-degrees of the points. It is shown that certain classes of these restricted digraphs are organized around a cyclic line, which cyclic line is better known to secular empiricists as the "in-crowd." Chapter 13 extends digraph theory by assigning a positive or negative sign to each directed line. It is pointed out that these signed digraphs can be used to represent relationships which are intrinsically positive or negative, e.g. love and hate, association and avoidance, and criteria for the "balance" of a digraph are introduced. This approach facilitates a mathematical analysis of, for instance, Heider's conception of balance in attitudinal or interpersonal structures. In the final chapter other extensions of digraph theory are considered, in particular, the applications to markov chains and social flow systems.

This book, written by two mathematicians and a psychologist, is a mathematical text designed to demonstrate the applicability of digraph theory to the social sciences. The reader interested in empirical research might have preferred an inversion of the author ratio, but he is likely to find that his task has been made easier by the presence at the end of the book of a glossary of the most important concepts and a list of the major theorems, and also by the high standard of the proofreading and layout. He is also likely to find that his tendency to use precise definitions in his treatment of structural phenomena is considerably reinforced.

E. A. C. Thomas.

Ruhleben: A Prison Camp Society. By J. Davidson Ketchum, with a Foreword and Postscript by Robert B. MacLeod. London: Oxford University Press. (University of Toronto Press). 1966. Pp. xxiii + 397. 48s.

J. Davidson Ketchum, a young musician studying in Berlin, was caught in Germany by the outbreak of the First World War and spent 4 years as internee together with

several thousand other British citizens. They were confined in a camp on the race course of the Berlin suburb of Ruhleben—which was not perhaps so ironically named, for the internees did in fact live fairly peacefully. Many interesting and creative persons were interned at Ruhleben (which must have made conditions easier, judging by the variety of entertainments and activities described) and a number of them wrote about it subsequently. For around 10 years after his repatriation, music remained Ketchum's main interest, but then—possibly because of his Ruhleben experience—he turned to psychology and sociology. This book was to have been his doctoral thesis, but at the time of his death in 1962 when he was Professor of Psychology at Toronto University, it was still not quite finished. A final valuable chapter by Professor Robert B. MacLeod brings the story up to the end of the war and tentatively sketches the sociological and psychological implications which unfortunately Ketchum did not have time to elaborate.

The material is indeed rich and unusual—Ketchum writes with both the meticulous accuracy of an anthropologist describing a strange society, and the sensitivity of an artist. The Ruhleben internees built a remarkable society out of the chaotic conditions in which they were thrown together. Ketchum maintains that "for this social achievement two factors were mainly responsible: the extraordinary nature of Ruhleben's population, and the challenging conditions of their internment. The prisoners were almost a cross-section of British society, from the manor house to the slum; scarcely a trade or profession was unrepresented. All were jammed together in a stable yard—company directors and seamen, concert musicians and factory workers, science professors and jockeys. Few had ever met previously; their only common bond was their British citizenship. The Germans, for their part, merely penned them up and left them to their own resources; no work was required of them, no recreation provided, no social structure imposed. As long as order was maintained they were free to pass the time as they wished, mingle together as they chose, and work out, if they could, a collective way of life that would make 4 years confinement bearable. The whole picture suggests an elaborate social experiment—careless, but undeniably intrigueing."

The book is cast in a roughly chronological form. It begins with a description of the formative first 12 months leading to the time when "a prolix outburst of activity and organization turned the community into a complex society." The next 3 years constituted a stage of relative stability, and then in 1918, when several hundred of the older men were repatriated, the society began to decline. But interwoven with the narrative are chapters devoted to such topics as leadership and status, entertainment, government and politics. All are told with great detail and delicacy, often with considerable humour and with a wealth of illustrative material drawn from a mass of documents which Ketchum was able to assemble, very largely diaries and reports of camp activities.

Despite the fact that this study has more the character of an anthropological than a modern social psychological study with its wealth of statistical information, the material is nevertheless so detailed as to make it possible to draw psychological conclusions. Professor MacLeod skilfully discusses the problem of morale and comes to the conclusion—as might have been expected—that a good deal of the society's internal solidarity resulted from opposition to an external threat. He continues however that "there is something a little more encouraging that we can learn from Ruhleben. Granted that the firmness of its social structure depended in no small measure on forces from without, symbolized by the enemy, the great outburst of creative activity bore little or no relation to the enemy. Ruhlebenites became creative when, having stabilized the minimal routines of living the consequently eliminated sheer survival as a goal, they found themselves free to pursue goals that were intrinsically challenging and worthwhile. The presence of the enemy, inasmuch as this meant limitation of freedom to choose alternatives, was actually a facilitating factor. The physical cage functioned in a very real sense as a liberator." Ketchum was also interested in the implications of Ruhleben behaviour for a theory of human nature. This, unfortunately, he never had the chance to work out systematically, but finding that so many men, although sexually deprived and frequently hungry, behaved as creative and co-operative members of society, made him impatient of some theories which would give over-simplified explanations of human behaviour as a direct response to environmental stimuli.

This is an extraordinary source book for the study of special communities. What it lacks in quantifiable material it makes up in graphic detail of both individuals and social activity. It may indeed, as Professor MacLeod believes, become a minor classic of social science.

ADAM CURLE.

Individual Differences. By Anne Anastasi. London and New York: Wiley. 1965. Pp. xiii + 301. 40s. cloth. 25s. paper.

This most recent of the "Perspectives in Psychology" publications is a collection of 31 articles and book-extracts spanning almost a century, from 1860 to 1962. The author has been modest in including only one article of her own in the collection; her main contribution lies in selecting the articles, in arranging them into a coherent pattern, and in supplementing them by an introductory chapter and connecting commentaries.

The book is not a long one, and restrictions have been inevitable. The individual differences in the title refer only to differences of ability, and almost exclusively to human ability: the one chapter on individual differences in animals, whose purpose is to illustrate the value of selective breeding experiments, has an air of not quite belonging, of having been dragged in to fill an obvious gap in the human story.

The chosen articles centre round four main themes, the measurement of ability, the nature of intelligence, the relative contributions of heredity and environment to intelligence, and creativity. Most of the contributions, if not all, are very well-known and many of them are classics. Galton, Pearson, Spearman, Thurstone, Thorndike, Burt, Cattell each makes his expected appearance: the problem in these cases has clearly been one of deciding which excerpt to choose for inclusion. Inevitably, a good deal of cutting has been necessary, and the result, in Galton's case at least, is at times a rather jerky, disconnected narrative which is at variance with his normally smooth fluid style.

There is, then, nothing very new in the content of the book. What the author has set out to do is to put the contributions of these authors in historical perspective, and to trace the change in opinion from the dogmatic hereditarianism of the 1900's epitomized by Pearson and Thorndike, to the more optimistic environmentalism of the last two decades. Her own article and her editorial commentaries underline this change, and stress the real problem: how, with the knowledge which we now have both of genetic and environmental influences on ability, to ensure the optimal conditions for the development of the individual.

The book is a useful one for a student trying to find his way through the mass of literature on individual differences. It is not, and it does not attempt to be, a comprehensive survey of the field: what it does do is to provide the student with a manageable framework into which he can fit many more contributions than are even mentioned by Anastasi.

ELIZABETH D. FRASER.

Cognition and Thought: An Information-Processing Approach. By Walter R. Reitman. London and New York: Wiley. 1965. Pp. xiii + 312. 60s.

"My main intent is not to review or survey, but rather to present a treatment of selected topics in cognition and thought, to some extent along the lines that Hebb developed, but making use of the very powerful information-processing concepts and methods developed over the past decade."

The information-processing approach, we are told, provides a concise and unambiguous language within which to theorize. This language may be used informally for the communication of ideas and as a tool for speculation. Alternatively, it enables theories to be formalized as computer programs so that their implications become explicit. How are such program-theories tested empirically? Current procedures are crude versions of Turing's test: the program's performance is compared with the behaviour and protocols of human subjects. We may simply count the number of "errors" the program makes, or, in addition, set the program "back on the track" whenever it diverges from the subject's performance. The author outflanks the deficiencies of these and other tests. He insists that the *testing* of predictions is inappropriate, a program-theory should be continuously *modified* in the light of empirical data. He attacks the notion that theories must be tied to operational definitions, and minimizes the predictive power of many conventional psychological theories, regarding them—like information-processing theories—as repositories of key concepts.

The main features of a list-structure information-processing language, IPL–V, are described, with details in the appendix. This language is utilized to introduce a frame-word for models of cognitive structure, procedures for handling concepts and rules, and a low level analysis of problem-solving. Newell, Shaw, and Simon's program, to which a chapter is devoted, simulates the solving of well-defined problems, e.g. proving theorems in the propositional calculus. But, greater difficulties arise with ill-defined problems. Their solution necessitates the closing of "open" constraints, hence the

multiplicity of possible solutions. This is demonstrated in the lengthy analysis of a composer's remarks whilst writing a fugue. We are given not a fragment of the resulting score—for some, a rather frustrating economy!

An interesting chapter describes the author's program, Argus, which solves simple analogy problems. Its genealogy is evidently *by* Hebb *out of* Newell, Shaw, and Simon. An innovation is the *active* semantic network that constitutes its memory. The elements of this network represent entities in the real world. They are connected by various relations so that the "firing" of one element may lead to the "firing" of others. An element's state of readiness changes with both the passage of time and the activity of the network itself. This activity occurs in parallel more or less independently of the serial organization of the problem-solving strategies. So, "forgetting," "distractability," and the ability to "pick up the threads" again, are all simulated. The author rightly claims that this network could feature in many diverse programs that simulate cognition.

Finally, there is an analysis of comprehension and the conversion of information into an internal symbolic representation. Several programs that simulate aspects of these processes are described, and the author's ideas make an interesting contrast with other current theories of semantics.

The merit of the book is that it requires no previous knowledge of computer techniques. It is a good introduction to their use in cognitive theory. But its bias away from empirical predictions and findings diminishes its general psychological interest. The author writes clearly if somewhat redundantly. He suffers from "programmer's syndrome," i.e. a predilection for turning statements into imperatives ("Note that . . .") and for excessive signposting to the contents of other chapters.

P. N. JOHNSON-LAIRD.

Mental Deficiency: The Changing Outlook. Edited by Ann M. Clarke and A. D. B. Clarke. London: Methuen. 2nd Edition. 1965. Pp. xxi + 596. 63s.

A number of American books are available on the psychology of mental deficiency but, in England, there is only one comprehensive manual and that is the one compiled by the Clarkes. Their first edition, published in 1958, had great virtues but a weakness lay in bridging the gap between the medical and educational aspects of the subject. In the new edition this disadvantage has been removed by an excellent clinical summary, in Chapter VI, of present day knowledge on the genetical and environmental causes of mental subnormality. There is still a fair amount of repetition in the discussions on problems of incidence in the community and on the principles of training. This is presumably inevitable in a book to which eight writers contribute and which aims to cover a somewhat limited field thoroughly. However, every chapter is well written and gives a reliable, objective and well documented account of the aspect with which it is principally concerned.

There are one or two minor criticisms which occurred to the reviewer. In Chapter II, which concerns the distribution of intelligence in the population, the extensive survey by Roberts, Norman and Griffiths is not mentioned nor is the Scottish Survey seriously discussed here or in Chapter III. On p. 96, the expected hereditary likeness of first cousins, expressed by a correlation coefficient, seems to be indicated as 0·25 instead of 0·125. The tendency for psychological peculiarities to accompany specific physical types, such as mongolism, is considered, on p. 178, to be imaginary; but this may be because psychologists have so far failed to devise tests which describe the characteristic behaviour which is clinically detectable.

Chapter V, on the effects of environment, provides interesting data and should make disturbing reading for anyone who believed that intelligence, as commonly understood, was an entirely inborn and unalterable trait. The discussion, in Chapter VIII, on ability to learn in the mentally deficient is full of information of practical value. The last chapter, on adoption, gives a sympathetic and sensible account of numerous difficult questions which constantly require answers in this field.

Altogether the book is to be strongly recommended for those engaged in any branch of work concerned with the mentally handicapped, not only psychologists and teachers to whom it is primarily directed.

L. S. PENROSE.

Handbook of Mental Retardation Syndromes. By Charles H. Carter. Springfield, Illinois: Thomas. 1966. Pp. xiii + 168. $8.00.

This book catalogues the variety of pathological conditions of the brain, which may impair intellectual development in children. It is a strictly medical text, and in parts

relies heavily on medical terminology. The presentation is humdrum, and the style graceless. Nevertheless, it does gather together information widely scattered in the medical literature, and in this respect it may be of some use. M. KINSBOURNE.

The Biosocial Basis of Mental Retardation. Edited by Sonia F. Osler and Robert E. Cooke. London: Oxford University Press (Johns Hopkins Press). 1966. Pp. xiv + 151. 40s.

This non-book follows a current fashion. About half a dozen unconnected papers are bound together, provided with stiff covers and given an overall title, in the hope that, somehow, a coherent unity will emerge. It rarely does, and this present effort confirms this. The editors have gone even further; through the use of the term "biosocial" they seem to think they have created a new discipline. Unfortunately the contributors stick to their orthodox sociological or biological procedures, and "biosociology" remains confined to the title. Similarly, statements that the importance of the book lies in its assessment of qualitative differences within the field of mental retardation never progress beyond the blurb and the introduction. None of the authors ever mentions the subject.

Instead, following a sober chapter on the extent of the problem, there is one on the nature of intelligence, excellent for undergraduates who will have to write an examination answer on the subject, but of no other conceivable use. Then comes an informative chapter on perinatal factors associated with subnormality and a clear account of the way in which early deprivation of photic stimulation affects subsequent behaviour of monkeys, but this seems irrelevant to human mental retardation. Harlow contributes a chapter on induced mental and social deprivation in Rhesus monkeys which is entertainingly written, but the material is readily accessible elsewhere. Harlow's final offhand generalizations are distressing. The deprived monkeys showed no indication of intellectual loss but he wished to call them "autistic," although autistic children usually *do* show some degree of intellectual impairment. Therefore, argues Harlow, it follows that the autistic child does not acquire speech "because he isolates himself socially," the retardation being secondary. Statements of this kind do not arise from his data, do not fit with observations on humans and can be dangerously misleading. The book is concluded by two chapters on psychological investigations which are well known and add little to previous accounts.

The book would have served a useful purpose if it stimulated the formation of an association to protect the memory of John F. Kennedy. Perhaps it is impossible to prevent airports and missile launching sites being named after him but surely there is no need to dedicate to him indifferent compilations on mental retardation as well.

BEATE HERMELIN.

Prediction of Response to Pharmacotherapy. Edited by J. R. Wittenborn and P. R. A. May. Springfield, Illinois: Thomas. 1966. Pp. xii + 231. $10.00.

This book presents the proceedings of a Workshop on Prediction held by the American College of Neuropsychopharmacology in 1964. There are 11 contributions by authorities including Cole, Wittenborn and Lehmann. The orientation is mainly clinical. One chapter by Lehmann presents a very useful account of the many reasons for individual differences in reaction to drugs. This is followed by four papers on the problems of (*a*) how does one choose the right drug from many similar for this patient? (*b*) how can one classify psychiatric patients on some basis more useful than the classical clinical differentiation? (*c*) can one select from various previous pretreatment factors any which can be used to predict response to a particular agent? Under (*b*) explorations are made of various approaches based on personality theory, Gestalt principles and various statistical manoeuvres that will be of general interest. Next, there are two chapters discussing the relevance of animal research to these clinical problems, and the book ends with comment by a clinical psychiatrist, Dr. Klerman, and a summary by Dr. May.

The book presents an admirable picture of the sophisticated thinking current in American psychopharmacological circles upon how to improve clinical effectiveness in psychiatry. This reviewer is in sympathy with Dr. Klerman's opinion that a "reaction-type" approach to psychiatric illness should ideally give way to an aetiological approach expressed in biochemical terms. However, he is forced to conclude that since we know hardly anything about this at present "our best predictions are derived from behavioural and psychological or social measures. Future research will have to establish physiological or biochemical variables of predictive utility."

J. R. SMYTHIES.

Perspectives in Psychopathology: Readings in Abnormal Psychology. Edited by James O. Palmer and Michael J. Goldstein. London: Oxford University Press. 1966. Pp. xiii + 434. 32s.

The aim of this book is to provide background reading for an undergraduate course in psychopathology. The editors have selected nearly two dozen previously published papers, classified them roughly as experimental, biological, social, clinical and actuarial and provided linking comments. Some of the papers are well known and excellent individually. These include the critique of twin studies of schizophrenia (Jackson), the Bethesda studies on psychotropic drugs, biochemical theories of schizophrenia (Kety), and schizophrenic patients and their sibs (Lidz *et al.*). The book taken as a whole however lacks unity and the elementary level of the editor's comments is likely to irritate rather than stimulate English undergraduates. SIDNEY CROWN.

Hypnosis. By L. Chertok. Translated by D. Graham. Oxford: Pergamon Press. 1966. Pp. xv + 176. 45s.

The neglect and opposition that hypnosis has suffered in France, where so much of the early work was done, has prompted Dr. Chertok to review its historical development and present status in the hope that it may in the future play a part not only in therapy but also in the development of theories of mental functioning. The clinical section is disappointing: the few case-histories are unrevealing, and there is no guide whatsoever to the frequency and degree of improvement to be expected in various disorders. The bogy of symptom-substitution and of development of psychosis following removal of hysterical symptoms by suggestion is again raised, in spite of the evidence (e.g. Carter, 1949, *Brit. med. J.*, i, 1076) against this belief. The section on techniques of induction is informative, but useless to the novice, who will not get from this account the only thing he really needs, namely absolute confidence of success in his own first attempt. The most interesting section is on the historical aspects, where the author throws light on many little-known episodes. Nevertheless in spite of these criticisms this volume is attractively written (and translated), and is worth attention for its moderated enthusiasm and its resolve to integrate hypnosis into the field of psychiatry. R. T. C. PRATT.

Psychology: An Introduction to a Behavioral Science. By H. C. Lindgren, D. Byrne and L. Petrinovich. New York and London: Wiley. 2nd Edition. 1966. Pp. xv + 560. 60s.

The first edition of this textbook was published in 1961. It is not clear what specific changes have been made in this one, apart from bringing the references up to date. The pattern, however, follows closely that of several other recent texts produced for undergraduate consumption in the States, and reflects what might be called an attenuated Behaviourism. Among topics of comparatively recent interest that find inclusion are "creativity" and psychopharmacology, though there is nothing about information theory and very little about language. Hull still rates two references but Eysenck achieves none, a distinction he shares with Freud. Apart, indeed, from the work of Piaget and his school, there is virtually no reference to any work done outside the United States. This might seem a little odd when a whole chapter is devoted to "Psychology Applied to Problems of International Behavior." Perhaps one way to get this going would be to encourage psychology students to take a look at works in languages other than their own, not excluding English. O. L. ZANGWILL.

Basic Contributions to Psychology: Readings. Edited by Robert L. Wrenn. Belmont, California: Wadsworth. London: Prentice-Hall International. 1966. Pp. xi + 308. 24s.

From a short-list of 93 references which he found to be common to at least four of the five 1st year texts which he analysed the editor R. L. Wrenn selected this strange collection of 48 readings. Of these, 10 first appeared in the *American Psychologist*, 20 in other journals and 18 were culled from books, most of which are well known and readily available. Eighteen of the articles were first published before 1930 and 23 after 1950.

Wrenn has marshalled his 48 pieces, 30 of which he has abridged most inexpertly, into 10 chapters. For each of these he has written a short introduction which in most cases is embarrassingly bad on all counts. The net result is a worthless book selling at the price of half a bottle of whisky. J. R. SYMONS.

Reward and Punishment. By Frank A. Logan and Allan R. Wagner. Boston: Allyn & Brown. 1965. Pp. viii + 119. $2.25.

This is one of an interesting new series of monographs ("Contemporary Topics in Experimental Psychology") now being published under the general editorship of James Deese and Leo J. Postman and primarily intended for undergraduates. In a foreword the editors explain the aim of the series—namely to discuss in each volume one rather specialized topic, and to do this in a way illustrating methods applicable to related topics.

In the present volume the topic is narrower than the title might imply since the authors deal almost exclusively with rat experiments and the theories based on them. "We have largely ignored the research on human verbal and motor learning. Our contention is that the same basic principles apply to all organisms that can learn, but whatever rewards and punishments occur in the human laboratory are currently difficult to identify and quantify. Accordingly the interested reader should view this book as a foundation for further study."

After a brief discussion of the Law of Effect, and of classical conditioning, the authors summarize a large number of experiments and theories about incentives. Hull, Guthrie, Miller, Mowrer and Spence, are amongst the theorists discussed.

In spite of the extreme degree of condensation required to bring together so much material in such a small space, the authors have produced a very readable book, which will serve as a useful introduction to Behaviour Theory. G. C. GRINDLEY.

Insight vs. Desensitization in Psychotherapy: An Experiment in Anxiety Reduction. By Gordon L. Paul. London: Oxford University Press (Stanford University Press). 1966. Pp. 148. 40s.

This brief monograph reports on a comparison of the efficacy of treatment based on an "insight" model of psychotherapy with that based on a "learning" model in reducing anxiety aroused by having to speak in public. The subjects were the 96 out of a group of 710 students enrolled in a public-speaking course at an American university, whose scores on personality and anxiety scales showed them to be the most disabled. The experimental design was a sophisticated one. The "learning" method—"modified systematic desensitization"—achieved 100 per cent. success, the "insight" method and an "attention-placebo" method, both 47 per cent., and "no" treatment, 17 per cent. The author has also provided a concise review of the recent literature and has some interesting things to say about the processes through which treatment produces benefit.

D. RUSSELL DAVIS.

Determinants of Infant Behaviour III. Edited by B. M. Foss. With a Foreword by John Bowlby. London: Methuen. 1965. Pp. xiii + 264. 50s.

This volume is based upon the proceedings of the 1963 meeting of the Tavistock Study Group on Mother-Infant Interaction. It brings together both animal and human studies bearing upon the central theme and includes some edited discussion. Professor Foss has some interesting things to say about theories of imitation and Professor Hinde is brief though to the point about monkey aunts. There is a long paper by Dr J. L. Gewirtz (held over from the 1961 meeting) on smiling in infants, but in view of the fact that this paper has also appeared in substantially the same form in a volume published by the Hebrew University of Jerusalem (reviewed in this *Journal*, 1966, **18**, 288), it might seem difficult to justify its tardy appearance here. O. L. ZANGWILL.

An Introduction to Psychopathology. Second Edition. By D. Russell Davis. London: Oxford University Press. 1966. Pp. x + 158. 16s.

This is an entirely rewritten and greatly shortened version of a book first published in 1957. Unfortunately, the compression is so severe that the book now conveys a didactic, at times almost authoritarian, tone, that accords ill with the author's well-known tolerance to controversial issues. Among new topics discussed in this edition are chromosome abnormalities, minimal brain injury at or before birth, and behaviour therapy. This book should be very helpful to students of psychology seeking a brief account of current issues in psychopathology, though it should be borne in mind that the author's strongly environmentalist approach is not everyone's cup of tea. O. L. ZANGWILL.

INDEX TO AUTHORS

THE BRITISH JOURNAL OF PSYCHOLOGY

Edited by ARTHUR SUMMERFIELD

Contents of Volume 57, Parts 1-2, 1966

REVUE CANADIENNE CANADIAN JOURNAL DE PSYCHOLOGIE OF PSYCHOLOGY

VOLUME 20, NO. 3 SEPTEMBER 1966

Published for the Canadian Psychological Association by
UNIVERSITY OF TORONTO PRESS

THE CANADIAN JOURNAL OF PSYCHOLOGY is published quarterly. Annual subscriptions, $6.00, single number $1.75.

Contributions. Original manuscripts and correspondence on editorial matters should be sent to:

The Editor, Canadian Journal of Psychology, McMaster University, Hamilton, Ontario.

Subscriptions to Journal. Orders and correspondence regarding subscriptions, changes of address, and purchase of back numbers shall be sent to:

Canadian Psychological Association,
The Secretary-Treasurer,
Suite 210, 225 Lisgar Street,
Ottawa 4, Ontario.